



Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec



Reconstructing the yield curve[☆]

Yan Liu^a, Jing Cynthia Wu^{b,c,*,*}

^a Purdue University, West Lafayette, IN 47905, USA

^b University of Notre Dame, Notre Dame, IN 46556, USA

^c National Bureau of Economic Research, Cambridge, MA 02138, USA



ARTICLE INFO

Article history:

Received 3 August 2020

Revised 20 December 2020

Accepted 21 December 2020

Available online 2 June 2021

JEL classification:

E43

G12

C14

Keywords:

Yield curve

Term structure of interest rates

Return forecasting regressions

Excess volatility

Non-parametric method

ABSTRACT

The constant maturity zero-coupon yield curve for the US Treasuries is one of the most studied datasets. We construct a new yield curve using a non-parametric kernel-smoothing method with a novel adaptive bandwidth specifically designed to fit the Treasury yields. Our curve is globally smooth while still capturing important local variation. Economically, we show that applying our data leads to different conclusions from using the leading alternative data of Gürkaynak et al. (2007) (GSW) when we repeat two popular studies of Cochrane and Piazzesi (2005) and Giglio and Kelly (2018). Statistically, we show our dataset preserves information in the raw data and has much smaller pricing errors than GSW. Our new yield curve is maintained and updated online, complemented by bandwidths that summarize information content in the raw data.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The constant-maturity zero-coupon Treasury yield is a critical dataset for researchers in macroeconomics and finance for studying the term structure of interest rates (e.g.,

Ang and Piazzesi, 2003; Hamilton and Wu, 2012; Diebold and Rudebusch, 2013; Wu and Xia, 2016), estimating return forecasting regressions (e.g., Fama, 1984; Cochrane and Piazzesi, 2005), examining international yield curves and exchange rates (e.g., Lustig et al., 2019; Chernov and Creal, 2020), studying bond term premia (e.g., Rudebusch and Swanson, 2012; Creal and Wu, 2020), analyzing monetary policy (e.g., Bernanke and Reinhart, 2004; Swanson and Williams, 2014), and pricing other assets and derivatives (e.g., Hull and White, 1990; Jarrow and Yildirim, 2003). We construct a new dataset of zero-coupon yields that better represent information in the raw data. We make it available to researchers and will update it regularly: <https://sites.google.com/view/jingcynthiawu/yield-data>.¹

The most popular zero-coupon Treasury yield curve datasets are Fama and Bliss (1987) and Gürkaynak et al.

[☆] We appreciate comments from an anonymous referee, Michael Bauer, Anna Cieslak, John Cochrane, Drew Creal, Richard Crump, Jeroen Dalderop, Ian Dew-Becker, Stefano Giglio, Jim Hamilton, Bryan Kelly, Ralph Koijen, Monika Piazzesi, Bill Schwert (the editor), and Eric Swanson, as well as conference and seminar participants at the 5th Annual UWO Conference on Financial Econometrics and Risk Management, the 2019 Asia Meeting of the Econometric Society, and the 2020 ASSA Annual Meeting (San Diego). We are especially grateful to Cam Harvey for comments on several versions of this paper. Cynthia Wu gratefully acknowledges support from the Institute for Scholarship in the Liberal Arts, College of Arts and Letters, University of Notre Dame. Our yield curve data are available at <https://sites.google.com/view/jingcynthiawu/yield-data>.

* Corresponding author at: University of Notre Dame, Notre Dame, IN 46556, USA.

E-mail addresses: liu2746@purdue.edu (Y. Liu), cynthia.wu@nd.edu (J.C. Wu).

¹ The raw CUSIP-level coupon-bearing Treasury bond data come from the CRSP Treasuries Time Series.

(2007) (GSW). However, both have their limitations. The Fama and Bliss (1987) data have limited maturities (1–5 years) and are only available at a monthly frequency.² Due to the unsmoothed nature, extending their data to more densely distributed or longer maturities is problematic. For example, Joslin et al. (2014) extend the Fama and Bliss (1987) data to maturities between 1 and 120 months. However, Cochrane (2015) shows Joslin et al. (2014) extended data feature large idiosyncratic measurement errors.

GSW exclude all underlying securities with less than three months to maturity and all Treasury bills. Therefore, by construction, the short end of their yield curve is extrapolated and has large pricing errors. An imprecise short end propagates to longer maturities beyond the first couple of coupon payments.³ Second, their parametric method, which focuses on fitting the medium-term maturities, also makes the long end of their yield curve subject to extrapolation errors. Third, their parametric approach with a small degree of freedom likely misses local information, which could be economically meaningful.

Our contribution is to construct a new zero-coupon yield curve that better captures the variation of the raw data, while it is as easy to use as the Fama and Bliss (1987) and GSW datasets, because the estimation step from the raw unbalanced coupon bond prices to balanced constant-maturity yields has already been completed. Our new yield curve data overcome the drawbacks of both Fama and Bliss (1987) and GSW.⁴

We construct our dataset with a kernel-smoothing method. Our dataset is at the daily frequency covering maturities of 1, 2,..., 360 months. We study its economic implications by revisiting the influential paper of Cochrane and Piazzesi (2005) (CP) for bond return forecasting regression, and the recent work on excess volatility of long-term bond prices by Giglio and Kelly (2018) (GK).

For the CP return forecasting regressions, our data are able to produce a robust loading pattern over the five forward rates, as in CP. The same pattern holds for bonds with different maturities and over different sample periods. This result is consistent with CP's main conclusion that one return forecasting factor, which is a linear com-

bination of forward rates, predicts excess returns. By contrast, estimates based on the GSW data do not produce a one-factor interpretation: the estimated loadings do not have a consistent pattern across maturity ranges or over time. Moreover, their loadings differ by an order of magnitude between CP's original sample period up to 2003, and the sample extended through 2019.

Next, we test the spanning hypothesis (whether the three yield factors are sufficient to predict bond returns) by regressing excess returns on the five principal components of the five forward rates. Using our yield curve, we find the fourth and fifth principal components have additional predictive power whether we base our conclusion on standard inference or the bootstrap procedure recently developed by Bauer and Hamilton (2018). This result is consistent with CP's conclusion as well as the literature that points to unspanned factors (e.g., Duffee, 2011). By contrast, with the GSW data, the higher-order principal components fail to show additional predictive power using CP's original sample period. Moreover, both the sign and the order of magnitude of the loadings change between CP's original sample period and the extended sample, indicating instability of the GSW yield curve. Overall, in the context of CP regressions, we show our data lead to economically different conclusions from those based on the GSW data: our data produce a robust loading pattern over maturities and time, and support unspanned yield factors.

For excess volatility on long-term bond prices, we calculate the variance-ratio statistic proposed by GK, which is used to compare the unrestricted variance of the log price of a long bond with its variance imposing the no-arbitrage restriction. For 20-, 25-, and 30-year bonds, we replicate GK's variance ratios of 1.19, 1.38, and 1.62, respectively, using the GSW data. However, GK find larger variance ratios (often above 2) across a variety of alternative asset classes. We show their smaller estimates for Treasury securities are likely driven by the GSW data. Using our data instead, the corresponding variance ratios become 1.63, 2.02, and 2.37, respectively, which is consistent with their overall conclusion, and significantly strengthens their Treasury results. Our data thus lead to quantitatively different results than those based on the GSW data.

Economically, what explains the differences between our data and GSW's through the lens of Cochrane and Piazzesi (2005) and Giglio and Kelly (2018)? Could it be microstructure noise? Cochrane and Piazzesi (2005) have spurred a large literature that rationalizes results of CP regressions from a risk perspective (e.g., Duffee, 2011; Chernov and Mueller, 2012; Joslin et al., 2014). Giglio and Kelly (2018) conduct extensive analysis to rule out microstructure noise as an explanation for their results. Therefore, a more plausible explanation is that our data better captures underlying signals about risk factors than GSW. Statistically, the main difference is that we use a non-parametric kernel-smoothing method based on Linton et al. (2001), whereas GSW apply Svensson (1994) extension of Nelson and Siegel (1987) parametric function. The non-parametric approach allows us to generate a globally smooth yield curve across maturities while still capturing economically meaningful local variation.

² Their data are available from CRSP Treasuries.

³ We use an example to illustrate. For simplicity, let us assume coupon/maturity calendars are exactly multiples of six months. A 1.5-year bond has two coupon payments at 0.5 and 1 year. An imprecise short end up to a year price both coupons with errors, leading to an imprecise discount rate at 1.5 years. Moving forward, the imprecise discount curve up to 1.5 years jeopardizes the discount rate at 2 years, and so on. Therefore, discount rates at all horizons are thus affected.

⁴ The choice of the yield curve data inevitably depends on the economic question we ask. For many macroeconomic applications, the choice of the dataset is likely inconsequential. However, for many other applications, different datasets may have significantly different implications. Prominent examples include return forecasting regressions and unspanned yield curve factors (e.g., Cochrane and Piazzesi, 2005; Duffee, 2011). Cochrane and Piazzesi (2009) show using the GSW dataset reduces return predictability because the yield curve is too smooth to capture information in higher-order principal components. Gürkaynak et al. (2010) made a similar point. Moreover, GSW is not suitable for pricing derivatives or any applications that require the short-term yields.

We propose a novel approach to adaptively select bandwidths, which are the key to non-parametric methods. This procedure is specifically designed for the Treasury yield curve. The bandwidth is inversely related to the number of observed Treasury securities around a given maturity. The bandwidth is smaller when we have more data. Conversely, parametric methods such as GSW have a fixed degree of freedom across all maturities. Our adaptive bandwidth selection allows us to keep securities at the short end of the yield curve including Treasury bills, which we find contain important information in disciplining the overall behavior of the yield curve. In general, the flexibility of our non-parametric approach enables our dataset to represent information in the raw data, not only for the medium, but more importantly, for the short and the long run.

Our online zero-coupon yield curve data are complemented by a bandwidth file, which captures how much information is in the raw data. In general, the short term is associated with the smallest bandwidth, implying ample observations. By contrast, the bandwidth at maturities longer than 10 years is often large, due to intermittent issuance. Although a popular choice in the literature, the 30-year yield sometimes pools information from bonds with maturities that are 10 years away, regardless of whether the data are constructed with parametric or non-parametric methods, even for the post-1990 sample.⁵ We recommend researchers use the bandwidth data as additional information to assess the quality of the zero-coupon yield curve and the availability of the raw coupon-bearing Treasury securities.

Another methodological contribution is a sequential method to systematically detect and delete outliers in the raw data as opposed to ad hoc methods used, for example, by [Fama and Bliss \(1987\)](#) and [Gürkaynak et al. \(2007\)](#). Our algorithmic approach is transparent and replicable.

Having discussed how we construct the yield curve and its economic implications, we next probe its statistical performance, both in sample and out of sample. We find that for maturities of less than one year, GSW generate large and sometimes extreme pricing errors, with the mean absolute error (in annualized yield to maturity) as high as 7%, whereas our method reduces this number down to 0.2%. The difference is not only in the short end. The reduction in the average pricing error across all maturities is 55%.

The better in-sample performance of our yield curve is not an artifact of the more flexible non-parametric curve. We also find smaller pricing errors in two out-of-sample prediction exercises: one leave-one-out cross-sectional prediction and one time-series prediction. For example, in the leave-one-out exercise, the average reduction in the out-of-sample pricing error is 49% across maturity buckets. We also consider several alternative specifications of our model for robustness checks. First, we vary the weights we put on different maturity segments; specifically, we downweight the short end. Second, we experiment with discarding more outliers. Third, we vary the bandwidth choices. The common conclusion is that our

⁵ A few examples that use the 30-year yield from the GSW data are [Gürkaynak et al. \(2010\)](#), [Gilchrist and Zakravsek \(2012\)](#), and [Giglio and Kelly \(2018\)](#).

method outperforms GSW's with respect to all alternative specifications we consider.

Besides better statistical performance, we also demonstrate that a term structure model fitted to our dataset prices the raw coupon-bearing Treasury securities with smaller pricing errors than the one fitted to the GSW dataset. Most prominently, the pricing error for securities between 0 and 3 months is 15 bps using our dataset, whereas the corresponding number is 23 bps for the GSW dataset, which is over 50% larger.

The rest of the paper is organized as follows. In [Section 2](#), we describe the non-parametric kernel-smoothing method. In [Section 3](#), we discuss the adaptive bandwidth-selection procedure. In [Section 4](#), we provide details on our outlier-detection algorithm. In [Section 5](#), we study the economic implications of our new data, and in [Section 6](#), we focus on its statistical performance. We offer concluding remarks in the final section.

2. Kernel-smoothing method

The goal is to extract a zero-coupon yield curve $y(n)$ for any maturity $n \in \mathcal{N}$ from observed Treasury bills, notes, and bonds, many of which have coupon payments. For theory, we use the support $\mathcal{N} = (-\infty, +\infty)$. In our application, we make it $\mathcal{N} = \{1, 2, \dots, 360\}$ months.⁶

Estimation of the yield curve amounts to minimizing a weighted average of the distance between the fitted price and the observed price across all available bonds. The number of yields on the zero-coupon curve often exceeds the number of observations. To guarantee uniqueness and smoothness, one needs to impose additional constraints on the minimization problem. For example, [Nelson and Siegel \(1987\)](#) and [Svensson \(1994\)](#), which are the underlying methods GSW follow, assume a parametric functional form for the yield curve.⁷

Alternatively, we rely on a non-parametric method. The main advantage of a non-parametric framework is that the yield curve does not need to have the same functional form across all maturities. Typically, the short end of the yield curve has more local patterns, whereas longer-term yields are smoother. We design our non-parametric method with an adaptive bandwidth (in [Section 3](#)) to specifically target this feature of the yield curve. By contrast, parametric methods including the ones in the literature struggle to capture both features and need to make compromises.

Our framework builds on the work of [Linton et al. \(2001\)](#), who introduced a non-parametric kernel-smoothing approach in estimating the yield curve. In particular, the authors focus on the asymptotic distribution of the yield curve estimate when it is assumed to be locally linear.

⁶ For earlier years when relatively long-maturity bonds are not available, the support is $\mathcal{N} = \{1, 2, \dots, n\}$, where $n < 360$ is the maturity limit that we specify later on.

⁷ Another parametric approach to estimate the yield curve is to impose the no-arbitrage condition with a term structure model. For example, see [Fontaine and Garcia \(2012\)](#), [Andreasen et al. \(2019\)](#), and [Pancost \(2020\)](#).

Different from their paper, we focus on the empirical performance based on a finite sample of bonds. Specifically, our goal is to construct a smooth zero-coupon yield curve that best describes the raw data. We make the following methodological contributions. First, we propose a new method for bandwidth selection in Section 3 targeting the unique features of the Treasuries. Second, we provide yield estimates over a denser set of maturities than in the literature, namely, $\mathcal{N} = \{1, 2, \dots, 360\}$. Third, we derive analytical derivatives for the first-order conditions of the objective function to facilitate computation. Fourth, our objective function is weighted by durations of bonds, which follows the literature on fitting the yield curve parametrically and is new to the non-parametric literature.

2.1. Pricing error for a security

At a given point in time, suppose we focus on a generic bond.⁸ It is characterized by its observed price p , its sequence of cash flows $\{c_j\}_{j=1}^J$ including its principal, and the corresponding maturities $\{\nu_j\}_{j=1}^J$. Given $y(\nu_j)$, the implied bond price \hat{p} is:

$$\hat{p} = \sum_{j=1}^J c_j \exp(-y(\nu_j)\nu_j). \quad (1)$$

The goal is to extract the entire zero-coupon yield curve $y(n)$ for $n \in \mathcal{N}$ from observed bond prices. Note $n \in \mathcal{N}$ denotes a maturity on the constant-maturity zero-coupon yield curve, whereas ν_j denotes the maturity of the cash flow c_j . $\{\nu_j\}_{j=1}^J$ do not cover the entire support \mathcal{N} , nor is ν_j necessarily in \mathcal{N} . Therefore, we cannot obtain $y(n)$ by simply inverting (1).

Instead, we connect a given ν_j with an arbitrary $n_j \in \mathcal{N}$ by approximating $y(\nu_j)$ with $y(n_j)$ using a first-order Taylor expansion:

$$y(\nu_j) \approx y(n_j) + (\nu_j - n_j)y'(n_j), \quad (2)$$

where $y'(n_j)$ is the first derivative of the yield curve evaluated at n_j . Now, we can approximate the bond price in (1) using (2):

$$\begin{aligned} \hat{p}(n_1, n_2, \dots, n_J) \\ \approx \sum_{j=1}^J c_j \exp[-(y(n_j) + (\nu_j - n_j)y'(n_j))\nu_j], \end{aligned} \quad (3)$$

where each $y(\nu_j)$ for the cash flow c_j is approximated by an arbitrary point on the zero-coupon yield curve $y(n_j)$.

In general, n_j could be any maturity in \mathcal{N} . However, the closer n_j is to ν_j , the more information the j -th coupon payment provides on $y(n_j)$. To capture this idea, we use a normal kernel-weighting function:

$$\begin{aligned} K(n_j, \nu_j) &= K_h(\nu_j)(n_j - \nu_j) \\ &= \frac{1}{\sqrt{2\pi h(\nu_j)^2}} \exp\left[-\frac{(n_j - \nu_j)^2}{2h(\nu_j)^2}\right], \end{aligned} \quad (4)$$

where $h(\nu_j)$ is the bandwidth parameter or the standard deviation of the normal distribution. The weighting function has two features. First, given the bandwidth, the weight is higher when n_j is closer to ν_j . Second, the bandwidth $h(\nu_j)$ is a function of ν_j . This is essential for our application and allows us to pool information more locally around one maturity and more globally around another.

When $h(\nu_j)$ goes to zero, the cash flow c_j only provides information for $y(\nu_j)$, and does not provide any information for $y(n_j)$ when $n_j \neq \nu_j$. Therefore, a narrow bandwidth overweights information locally and tends to generate a non-smooth yield curve. On the other hand, when $h(\nu_j)$ approaches infinity, all maturities are weighted equally. Hence, a wide bandwidth pools information more globally but may generate yield curves that are overly smooth and lack local variation. We design our bandwidth to specifically target features of the Treasury yield curve; see details in Section 3.

Given the kernel weights, the kernel-weighted squared pricing error is:

$$\varepsilon = \int \dots \int (p - \hat{p}(n_1, n_2, \dots, n_J))^2 \prod_{j=1}^J K(n_j, \nu_j) dn_j, \quad (5)$$

where $\hat{p}(n_1, n_2, \dots, n_J)$ is defined in (3). Note we have $\int \dots \int \prod_{j=1}^J K(n_j, \nu_j) dn_j = 1$, and $K(n_j, \nu_j)$ is positive everywhere, which makes it an appropriate density function.

There are two reasons for our choice of the normal kernel. First, the fitting behavior is similar among different kernels, but the normal kernel has an advantage due to its analytical tractability (see, e.g., Wand and Jones, 1994). Second, in our framework, the continuous differentiability of the normal kernel allows us to derive the first-order conditions associated with (5) analytically, which greatly facilitates our estimation of a large dimensional yield curve (see Appendix A.1 for the derivation).⁹

2.2. Summarizing information across bonds

We have thus far constructed the kernel-weighted squared pricing error for a generic bond. To combine information from all available bonds at a given point in time, we need to add up the squared pricing errors across bonds. Suppose I bonds are available on a given day. Let the kernel-weighted squared pricing error for bond i be ε^i for $i = 1, \dots, I$, where ε^i is defined in (5).

The same discrepancy between the actual price and the fitted price has different implications for two bonds that have different maturity structures. For example, a \$1 pricing error is more pronounced for a short-term Treasury bill as opposed to a 10-year Treasury note. This difference can be captured by weighting ε^i with $1/D_i^2$, where D_i is bond i 's duration, defined as:

$$D = \sum_{j=1}^J \frac{\nu_j c_j \exp(-\nu_j \bar{y})}{p},$$

⁹ By contrast, alternative kernels such as the box kernel or the Epanechnikov kernel are not differentiable at the boundaries.

⁸ For brevity, we omit indicators for both time and bonds for now.

and the yield to maturity (YTM) \bar{y} is the constant discount rate that equates the present value of the bond's cash flows with its price:

$$p = \sum_{j=1}^J c_j \exp(-\nu_j \bar{y}). \quad (6)$$

The duration-weighted pricing error can be interpreted as the equal-weighted error in the yield space. Therefore, our objective function is:

$$S(y(\cdot), y'(\cdot)) = \sum_{i=1}^I \frac{1}{D_i^2} \cdot \mathcal{E}^i, \quad (7)$$

where $y(\cdot)$ is the yield function and $y'(\cdot)$ is its first derivative. Our goal is to minimize this objective function to obtain $y(n)$ and $y'(n)$ for all $n \in \mathcal{N}$.

This weighting scheme is new in the non-parametric framework. We consider using durations to weight bond prices to be important, because doing so allows us to put more weight on fitting the shorter end of the yield curve, which affects the coupon payments of bonds at all maturities. Several papers that estimate the yield curve parametrically have applied the same weighting scheme (e.g., Nelson and Siegel, 1987; GSW).

Minimizing the objective function (7) with respect to $y(\cdot)$ and $y'(\cdot)$ is a non-trivial optimization problem. The main issue is that the integral in (5) does not have a closed-form expression and needs to be approximated. Therefore, we need to choose a discrete support to facilitate computation. We choose $\mathcal{N} = \{1, 2, \dots, 360\}$ months, which is denser than in Jeffrey et al. (2006), for example. Our choice of a dense support in \mathcal{N} requires estimating a large number of parameters. To alleviate some numerical burden, we derive analytical derivatives of the first-order conditions for (7), which provides efficient and accurate estimates of the yield curve; see A.1. Once we have estimates of the yield curve over this discrete support, our framework permits a kernel-weighted interpolation scheme to provide estimates for maturities that are not in the support. Our choice of a dense support ensures the estimated yield curve is smooth over the entire maturity range.¹⁰ See details on estimation in A.1.

2.3. Model-implied bond price

The model-implied bond price is

$$\hat{p} = \int \dots \int \hat{p}(n_1, n_2, \dots, n_J) \prod_{j=1}^J K(n_j, \nu_j) dn_j. \quad (8)$$

Once we have the estimated $y(\cdot)$ and $y'(\cdot)$ over $\mathcal{N} = \{1, 2, \dots, 360\}$, we approximate this object with (see our derivation in A.1):

$$\hat{p} = \sum_{j=1}^J c_j \left(\frac{\sum_{n=1}^{360} K(n, \nu_j) \exp \left[- \left(y(n) + (\nu_j - n)y'(n) \right) \nu_j \right]}{\sum_{n=1}^{360} K(n, \nu_j)} \right). \quad (9)$$

¹⁰ Note that interpolation guarantees the estimated yield curve is always continuous. However, a less dense support may lead to kinks in the estimated yield curve, which makes the yield curve less smooth.

In empirical sections, we compute \hat{p} using (9), and then calculate its associated YTM.

3. Bandwidth

One main methodological contribution of our paper is to propose a bandwidth-selection method for the yield curve. The choice of bandwidth determines the smoothness of the estimated yield curve, which is crucial to generate a globally smooth yield curve while not missing important local variation. In Section 3.1, we propose our adaptive bandwidth-selection procedure for the yield curve, and in Section 3.2, we leverage the notion of bandwidth to summarize information content in the raw data.

3.1. Adaptive bandwidth selection procedure

We propose a data-driven approach for choosing bandwidths. We follow the basic idea of adaptive bandwidth selection in the literature on non-parametric estimators (see, e.g., Park and Marron, 1990; Fan and Gijbels, 1995; Ruppert et al., 1995). We are the first to apply an adaptive bandwidth-selection procedure to estimate the yield curve. Our specific choices are new to the literature.

For each ν that corresponds to a cash flow, we choose $h(\nu)$ such that N_0 bonds mature within the two-bandwidth interval around ν (i.e., $[\nu - 2h(\nu), \nu] \cup [\nu, \nu + 2h(\nu)]$). In our main analysis, we set N_0 at 8.¹¹ For a maturity segment with plenty of observations, the bandwidth $h(\nu)$ is small, and vice versa. To price this cash flow at ν , the relevant region in the zero-coupon yield curve is $n \in [\nu - 2h(\nu), \nu + 2h(\nu)]$, which covers 95% of probability weights.

In practice, observations are not equally spaced, and they are asymmetric around ν . For these reasons, we adapt our bandwidth selection procedure as follows. Let $N([\nu_a, \nu_b])$ denote the number of bonds whose maturities fall into the interval $[\nu_a, \nu_b]$. We first define the left-hand-side bandwidth at maturity ν (i.e., $h^l(\nu)$) as:

$$\begin{aligned} h^l(\nu) &= \frac{1}{2} \min b \\ \text{s.t. } N([\nu - b, \nu]) &\geq N_0/2. \end{aligned} \quad (10)$$

If no value of b satisfies $N([\nu - b, \nu]) \geq N_0/2$, we set $h^l(\nu)$ at $\nu/2$.

Similarly, we define the right-hand-side bandwidth at maturity ν (i.e., $h^r(\nu)$) as:

$$\begin{aligned} h^r(\nu) &= \frac{1}{2} \min b \\ \text{s.t. } N((\nu, \nu + b]) &\geq N_0/2, \end{aligned} \quad (11)$$

with $h^r(\nu) = \frac{1}{2}(n_{\max} - \nu)$ if $N((\nu, \nu + b]) \geq N_0/2$ cannot be satisfied for any b , where n_{\max} is the maximum maturity of the estimated yield curve.

Because the normal kernel is symmetric around ν , we consolidate $h^l(\nu)$ and $h^r(\nu)$ into one bandwidth:

$$h(\nu) = \min \{ \max \{ 3, h^l(\nu), h^r(\nu) \}, 120 \}, \quad (12)$$

¹¹ Our choice of $N_0 = 8$ is dictated by our out-of-sample forecasting results in Table C.3.

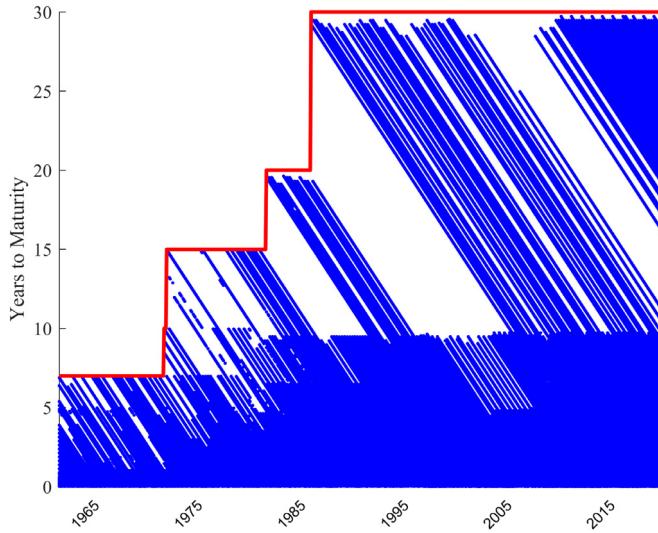


Fig. 1. Outstanding Treasury securities. Maturity distribution of outstanding securities, 1961–2019.

where three months is the minimum and 120 months (i.e., 10 years) is the maximum bandwidth we set for any maturity.

Calculating $h^l(\nu)$ and $h^r(\nu)$ separately guarantees we take information from both the left and the right sides of ν . This is important because the maturity distribution of outstanding Treasury securities on a given day often contains gaps, leading to asymmetry between $h^l(\nu)$ and $h^r(\nu)$; see Fig. 1. For example, suppose a 10-year gap is present in the maturity space: no bonds exist with maturities between $\nu_a = 120$ (i.e., 10 years) and $\nu_b = 240$ (i.e., 20 years). Also suppose a large number of bonds exist with maturities that fall just below $\nu_a = 120$, implying $h^l(\nu_a)$ is small (in particular, $h^l(\nu_a) \ll 60 = \frac{1}{2} \times (240 - 120)$). Now consider the bandwidth choice at ν_a . If we set the bandwidth $h(\nu_a)$ at $h^l(\nu_a)$, the bond price at maturity ν_a only provides information for the yield curve up to maturity $\nu_a + 2h^l(\nu_a)$,¹² leaving the majority of the yield curve between ν_a and ν_b undetermined. Our solution is to set the bandwidth at $h^r(\nu_a)$, which is the larger one between $h^l(\nu_a)$ and $h^r(\nu_a)$.

For shorter maturities, many observations contain potential micro-structure noise and liquidity issues. For a fixed $N_0 = 8$, the bandwidth of $\max\{h^l(\nu), h^r(\nu)\}$ tends to be small. For example, $\max\{h^l(\nu), h^r(\nu)\}$ is, on average, around 0.5 months at the maturity of three months. Such a small bandwidth tends to generate substantial local variation in the estimated yield curve, which may not reflect the underlying true yield curve. Therefore, our choice of a minimum bandwidth of three months allows us to pool information from maturities that are within half a year of ν to smooth out the estimated yield curve.

On the other hand, too large a bandwidth may bias the yield curve estimate because the Taylor expansion in (2) can be inaccurate. We therefore set the maximum bandwidth at 120 months. This maximum bandwidth only applies to long maturities where the data are sparse and have large gaps in the maturity distribution.

Fixing the number of local observations at N_0 allows us to pool roughly the same amount of local information to estimate the yield curve at each maturity. Another benefit is that it automatically adjusts for the total number of Treasury securities available at each date. When more bonds exist (as in the later part of our sample), bandwidths in general shrink, which allows us to better capture the local variation in the yield curve. Lastly, our bandwidth choice is controlled by only one parameter N_0 , which facilitates our out-of-sample forecasting exercise that chooses the optimal N_0 in Section 6.3.

3.2. Information content in the raw data

In this section, we leverage the notion of bandwidth to summarize the information content in the raw data. Different from (12), which is the bandwidth for each cash flow ν , we are now interested in the information contained at each maturity n on the zero-coupon yield curve. We propose using:

$$h(n) = \min\{h^l(n), h^r(n)\}, \quad (13)$$

where $h^l(n)$ and $h^r(n)$ are calculated using (10) and (11). Note that if b does not exist for (10), we set $h^l(n)$ at ∞ . The same applies to $h^r(n)$.

Why do we take the minimum instead of the maximum? We use the previous example with a 10-year gap in the maturity space between 10 years and 20 years to illustrate. For the bond with $\nu = 120$, $h^l(\nu) \ll h^r(\nu)$. But it needs to provide information to maturities within the

¹² This is only approximately true, because the normal kernel assigns a non-zero weight to any maturity. However, it assigns relatively large weights to observations that are within two bandwidths.

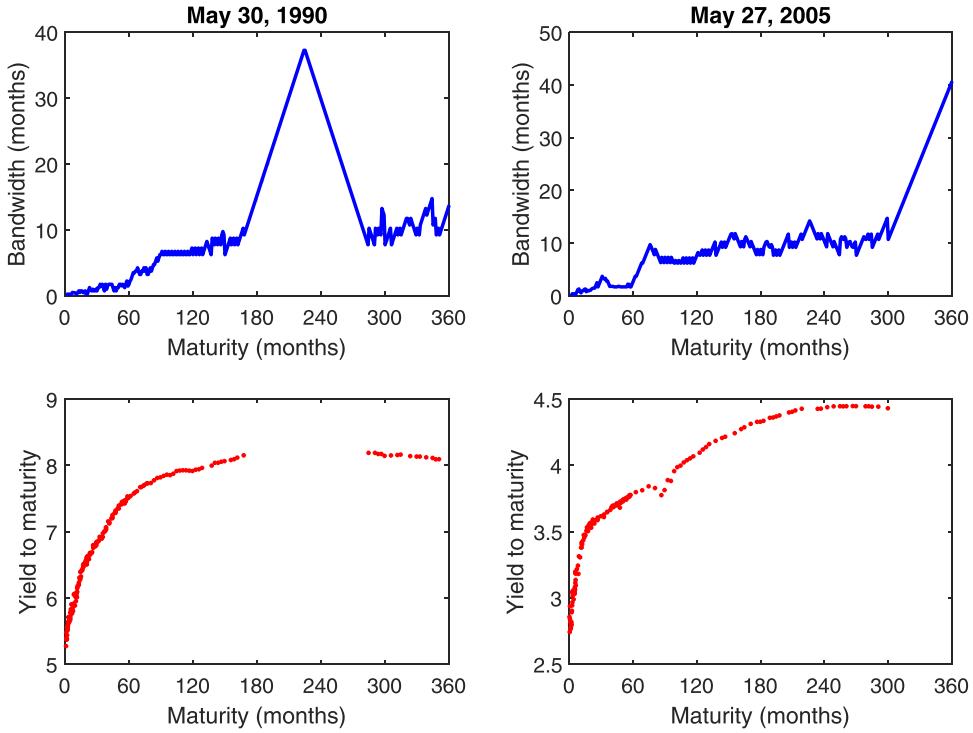


Fig. 2. Bandwidth on selected dates. We plot the cross-section of bandwidth for May 30, 1990 and May 27, 2005. The bandwidth is plotted in the top panels and YTM is plotted in the bottom panels.

gap $n \in (120, 240)$. This explains the maximum in (12). For $n = 120$ on the estimated constant-maturity zero-coupon yield curve, we still have $h^l(n) \ll h^r(n)$. However, the information we use to estimate the yield at $n = 120$ primarily comes from bonds on the left side, and $h^l(n)$ is small. Hence, we need to take the minimum instead.

We provide two examples in Fig. 2. The left panel is May 30, 1990, and the right panel is May 27, 2005. We plot bandwidths at the top. YTM is at the bottom and each dot corresponds to one outstanding security.¹³

For both dates, the bandwidth increases with maturity in general, indicating observations are more concentrated on the short end. For May 30, 1990, no outstanding securities have maturities between 180 and 296 months, which results in the spike in bandwidth. On May 27, 2005, the longest maturity is 300 months, and the bandwidth increases sharply after that.

Fig. 3 shows the time series of the bandwidth for various maturities, with the vertical bar indicating the beginning of 1990. Data on the short end are abundant, and the bandwidths for one, three, and six months are generally below 0.5 months.

The US Treasury does not always issue notes and bonds with longer maturities. In general, the bandwidths become smaller after 1990 for maturities longer than one year, but they remain large for maturities longer than 10 years. The 30-year yield is popular for studying the long end of the

yield curve in the literature. However, due to the intermittent issuance, even the post-1990 sample's bandwidth can get as large as 60 months, implying a lack of observations, and hence pulling information from bonds that are 10 years away.¹⁴

4. Raw data and outliers

4.1. Raw data

The raw CUSIP-level coupon-bearing Treasury bond data come from the CRSP Treasuries Time Series. For each bond, we observe the end-of-day bid and ask (and average) prices, maturity, coupon payments, and schedule, as well as other characteristics. The sample is from June 1961 to December 2019 at the daily frequency.

Fig. 1 summarizes the maturity structure for outstanding Treasury securities in our sample at the monthly frequency. We follow GSW and start including 10-year notes in August 1971, 15-year bonds in December 1971, 20-year bonds in July 1981, and 30-year bonds in November 1985. These dates mark the times when those maturities became more prevalent. We set the maximum maturity n_{\max} accordingly, which is marked by the horizontal lines in Fig. 1.

¹³ See Section 4.1 for details on the data.

¹⁴ Here we plot the bandwidth as defined in (13) to summarize information content in the raw data, and we do not impose a lower bound and upper bound as in (12).

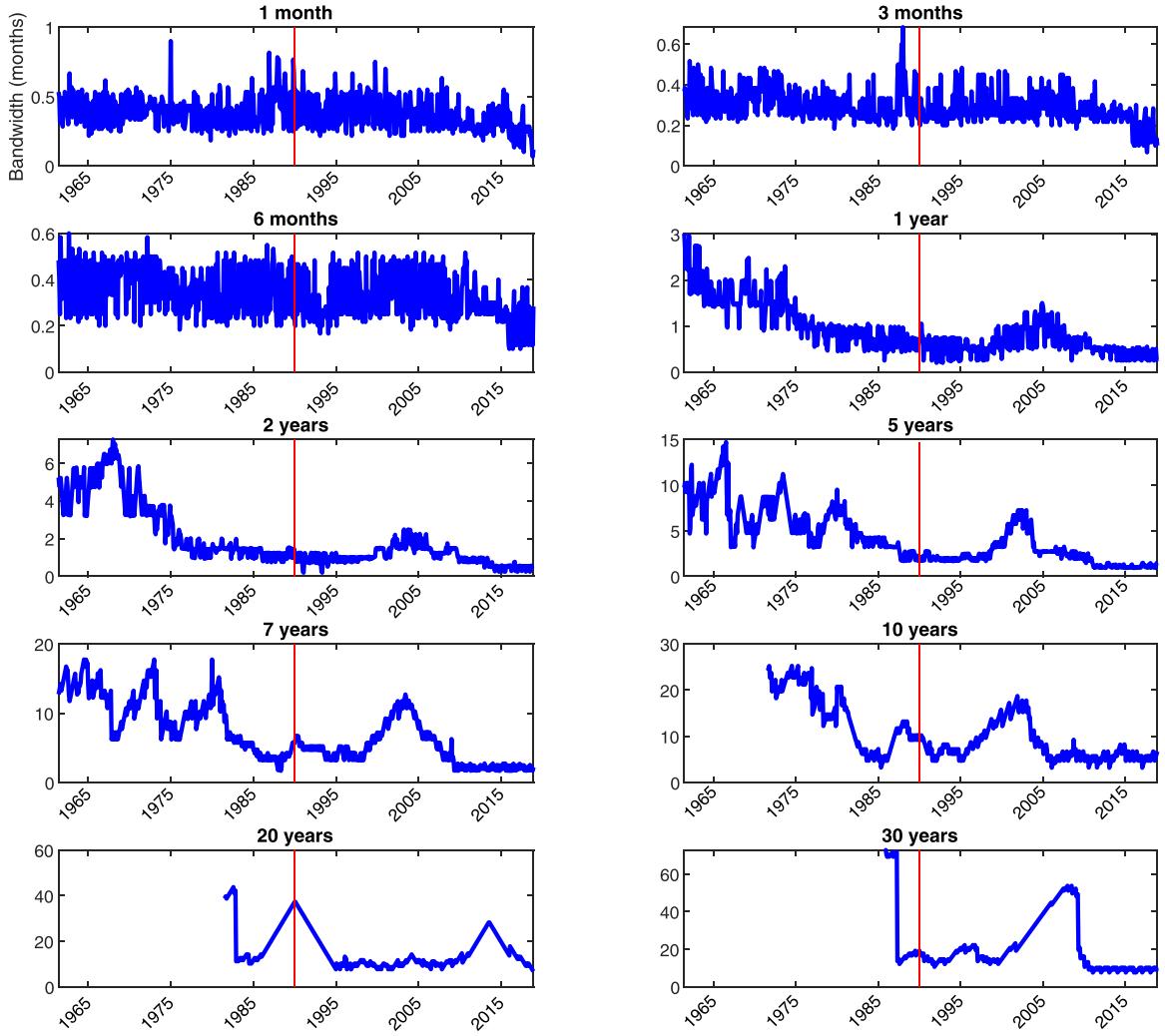


Fig. 3. Time series of bandwidth. We plot the time series of bandwidth over the entire sample from 1961 to 2019. The vertical bar marks the beginning of 1990.

4.2. Outliers

Our goal is to provide a single smooth yield curve that best represents the data. Therefore, we treat observations that do not belong to this curve as outliers. Our definition of outliers is similar to Fama and Bliss (1987) and Gürkaynak et al. (2007). Because our data preserve local variation in the yield curve, they can be used to address some questions related to trading frictions or microstructure issues. However, for researchers who are particularly interested in deviations from the smooth curve or the relationship between two curves (e.g., on-the-run vs. off-the-run issues), they may want to focus on the raw data.

In the literature on constructing a constant-maturity zero-coupon yield curve, outliers are usually deleted in an ad hoc fashion.¹⁵ We differ from the literature by taking

a more systematic approach. The nature of our approach makes it replicable for future research.

Note our outlier detection does not drive the performance difference between our yield curves and GSW's, as we document in Section 6. The reason is that we re-estimate GSW's yield curves based on the same filtered data, which facilitates a fair comparison.

Specifically, we take a three-step approach to drop outliers or inappropriate data from the perspective of fitting the yield curve. The first two steps are similar to the literature and ensure our estimation contains only plain vanilla bonds not severely affected by liquidity issues. The last step is our novel algorithmic detection procedure, which systematically deletes outliers.

The three steps are described as follows:

1. Only include fully taxable, non-callable, and non-flower bond issues (i.e., CRSP ITYPE equals 1, 2, 3, or 4).

¹⁵ For example, GSW state in Section 4, item (vi), "Other issues that we judgmentally exclude on an ad hoc basis."

This step ensures our sample does not include bonds with tax benefits and option-like features. [Fama and Bliss \(1987\)](#) apply the same filter.

2. *Exclude the two most recently issued securities with maturities of 2, 3, 4, 5, 7, 10, 20, and 30 years for securities issued in 1980 or later.*

This procedure follows GSW and aims to delete on-the-run (or first “off-the-run”) issues that often trade at a premium compared with other issues due to their liquidity and specialness.

3. *Sequentially delete outliers based on the fitted yield curve from the day before, where statistical cutoffs are computed using a segment of maturities.¹⁶*

This step is new. Our approach is algorithmic in nature (as opposed to the ad-hoc procedures in the literature), which makes it replicable.¹⁷

Steps 1 and 2 of our filtering procedure are similar to [Fama and Bliss \(1987\)](#) and GSW. The main difference from GSW is that we do not discard securities with shorter maturities or Treasury bills. We argue in [Section 3.2](#) that these securities contain important information, and in [Section 6](#), we show the importance of keeping them.

Three important features mark Step 3 of our filtering procedure. First, we use information from the previous day to adaptively drop outliers. This approach helps ensure internal consistency of the yield curve across days.¹⁸ Second, we compute summary statistics by maturity segments, and use them to determine outliers within each segment. This approach takes into account the differences in data quality across the maturity spectrum. Lastly, our statistical cutoffs adjust for time-varying data quality, allowing us to keep bond observations that are likely affected by microstructure noise during market stress or illiquidity.¹⁹ Our algorithm strikes a balance in dropping extreme observations and keeping information. We provide details in [A.2](#).

5. Economic implications of the new yield curve

With our newly constructed zero-coupon yield curve, we revisit three prominent studies using the US Treasury yield curve: the predictability of the bond risk premia of [Cochrane and Piazzesi \(2005\)](#) (CP), the excess volatility of

¹⁶ A few papers model on-the-run and off-the-run issues together, but they treat them differently. For example, whereas [Fontaine and Garcia \(2012\)](#) use additional liquidity factors to explain the wedge between on-the-run and off-the-run issues, [Pancost \(2020\)](#) treats their difference as a measurement error. Although we can, in principle, fit two yield curves, one on-the-run and one off-the-run, we choose to follow the consensus in the literature by focusing on off-the-run issues, because the number of on-the-run bonds is usually limited at each point in time.

¹⁷ Any outlier-detection algorithm has some subjective element to it. GSW exclude outliers on an ad hoc basis. [Fama and Bliss \(1987\)](#) drop observations when forward-rate reversal exceeds a pre-specified threshold level.

¹⁸ Although we do not drop outliers in an ad hoc manner, we do manually inspect the data with and without outliers to ensure our method does not systematically drop valid observations. We pay particular attention to coupon payment dates where CRSP Treasury data feature a discontinuity in bond prices and accrued interests.

¹⁹ Although we maintain the same 3.0-IQR rule in our outlier detection (see [A.2](#)), IQR becomes larger during market stress, allowing us to keep more observations.

the long-term bond prices of [Giglio and Kelly \(2018\)](#) (GK), and the Gaussian affine term structure model.²⁰

5.1. Cochrane-Piazzesi return forecasting regressions

In their seminal paper, CP find that although three yield factors explain the majority of the cross-sectional variation of the yield curve, an additional return forecasting factor, which is a linear combination of forward rates, predicts excess returns. This finding leads to a growing literature on the spanning hypothesis (whether the three yield factors are sufficient for predicting bond returns), which [Duffee \(2011\)](#) formalizes. For more references, see [Cooper and Priestley \(2009\)](#), [Ludvigson and Ng \(2009\)](#), [Greenwood and Vayanos \(2014\)](#), and [Cieslak and Povala \(2015\)](#).

We revisit CP’s analysis and study the economic consequences of using different underlying zero-coupon yield curves. We first repeat CP’s analysis using the same sample period and forward rates with the same maturities. Then, we extend their analysis along several dimensions.

First, we introduce some notation. Define the zero-coupon yield at t with a maturity of n as $y_t(n)$. The price of the n -year discount bond at time t relates to the zero-coupon yield as follows:

$$\log P_t(n) = -ny_t(n), \quad (14)$$

where n is maturity in years as in CP.

The forward rate with maturity n at time t is defined as the return for a loan starting at $t + n - 1$ and maturing at $t + n$:

$$f_t(n) = \log P_t(n - 1) - \log P_t(n). \quad (15)$$

The holding-period return of buying an n -period bond and selling it one year later is:

$$r_{t+1}(n) = \log P_{t+1}(n - 1) - \log P_t(n). \quad (16)$$

The excess return is:

$$rx_{t+1}(n) = r_{t+1}(n) - y_t(1), \quad (17)$$

where $y_t(1)$ is the one-year risk-free rate.

To repeat CP’s analysis, we run the following return forecasting regressions. The dependent variables are the excess returns of bonds with maturities of two to five years $rx_{t+1}(2), \dots, rx_{t+1}(5)$. The independent variables are the forward rates: $y_t(1), f_t(2), \dots, f_t(5)$. The regression has an intercept. The sample is CP’s original: monthly from 1964 to 2003.

In [Fig. 4](#), we plot the loadings (regression slope coefficients). Different lines represent maturities of excess returns (dependent variables) from two to five years. The x-axis denotes forward rates at different maturities (independent variables).

²⁰ We focus on the comparison of the conclusion drawn based on our data with that based on the GSW data. For most of our analyses in [Section 5](#) and [Section 6](#), we do not compare with the [Fama and Bliss \(1987\)](#) data because of their limited maturities (in total, five maturities available from one to five years) and frequency (monthly only). The only exception is CP, where we present results using ([Fama and Bliss, 1987](#)) in the Online Appendix.

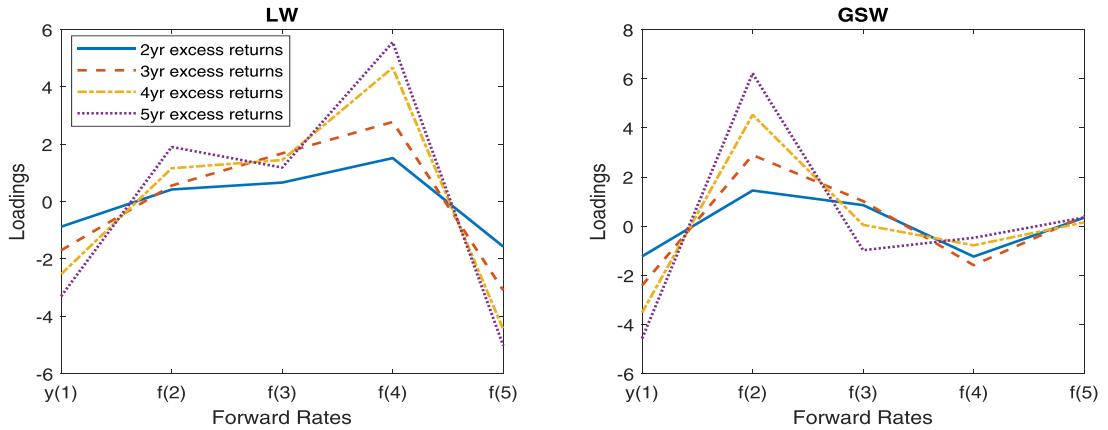


Fig. 4. CP loadings of individual excess returns. x-axis: independent variables $y_t(1), f_t(2), \dots, f_t(5)$. Different lines represent excess returns with different maturities $rX_{t+1}(2), \dots, rX_{t+1}(5)$. We use CP's original sample from 1964 to 2003.

In the left panel, we plot the loadings using our data. The five lines display the same “M” shape. CP highlight one single return forecasting factor that predicts all excess returns. Although we do not have the tent shape as in CP, the same “M” shape across different maturities is consistent with CP’s main conclusion of one return forecasting factor.²¹ By contrast, the patterns of loadings in the right panel using the GSW data are different across maturities. In particular, the loading on $f_t(3)$ is higher than $f_t(4)$ for the excess returns of two- to four-year bonds. But this pattern is reversed for the five-year bond.

Next, we summarize results for the average excess returns $\bar{r}X_{t+1}^{(2 \rightarrow 5)} = \frac{1}{4} \sum_{n=2}^5 rX_{t+1}(n)$. We take the average of the four lines in Fig. 4 to compute the average loading between two and five years, shown in the solid lines in the top panels of Fig. 5. We then extend the sample to the current period and show the loadings for the 1964–2019 sample using dashed lines. We also extend the dependent variables to longer maturities, and plot the loadings of 6- to 10-year bonds $\bar{r}X_{t+1}^{(6 \rightarrow 10)} = \frac{1}{5} \sum_{n=6}^{10} rX_{t+1}(n)$ in the middle panels, and 11- to 15-year bonds $\bar{r}X_{t+1}^{(11 \rightarrow 15)} = \frac{1}{5} \sum_{n=11}^{15} rX_{t+1}(n)$ in the bottom panels.

The same “M” shape preserves remarkably well with our dataset in the left three panels, which implies one single return forecasting factor prevails across different sample periods and maturities. That is not the case with the GSW data. Among the six lines in the right panels, the shape changes over time and across maturities. For example, in the bottom-right panel, the loadings between the

1964–2003 sample and the 1964–2019 sample are negatively correlated. Moreover, the loadings differ by an order of magnitude between the two samples: the largest loading (in absolute value) is 25.4 for the 1964–2003 sample but increases to 111.4 for the 1964–2019 sample.

The left panel of Table 1 reports the loadings of regressing the average bond excess returns between two and five years on $y_t(1), f_t(2), \dots, f_t(5)$. The loadings coincide with the top panels of Fig. 5.

The return forecasting regression has overlapping observations. To adjust for them, we make statistical inference using two alternative methods. First, we follow CP and calculate t -statistics and p -values using Newey-West standard errors with 18 lags. Second, we use the bootstrap procedure recently developed by Bauer and Hamilton (2018) (BH), and compute the 5% critical values based on the bootstrap distributions and the corresponding p -values.²²

We note that besides $y_t(1)$ (the log one-year bond yield), the other forward rates are never significant with the GSW data. With our data, $f_t(4)$ and $f_t(5)$ are highly significant using both Newey-West and BH bootstrap-based p -values. This result is consistent with the stable “M” shape loadings on forward rates in Figs. 4 and 5. Moreover, our R^2 s are in the same ballpark as the ones using the Fama and Bliss (1987) data (see the Online Appendix), whereas GSW produce smaller R^2 s.

We next conduct a test for the spanning hypothesis by regressing excess returns on the five principal components ($PC_1 - PC_5$) of the five forward rates in the right panel of Table 1. The null hypothesis for the spanning hypothesis is that the loadings on PC_4 and PC_5 are jointly zero. We report the F -statistics and the associated p -values, as well as BH’s 5% critical values based on the bootstrap distributions and the corresponding p -values.

Using our data, PC_4 and PC_5 are jointly significant in both sample periods with both standard and BH p -values.

²¹ We replicate CP’s tent shape using the Fama and Bliss (1987) data (see the Online Appendix). The main driver between their tent shape and our M shape is multicollinearity: $f_t(2), f_t(3), f_t(4)$ are highly correlated, with correlations between 0.97 and 0.98 in both ours and the Fama and Bliss (1987) data. The fact that multicollinearity drives the tent shape away is also found by Hodrick and Tomunen (2018) in their international study. Moreover, it also changes the tent shape by simply extending the Fama and Bliss (1987) data (see the Online Appendix). However, two main conclusions are consistent between the Fama and Bliss (1987) and our data. First, both imply higher loadings for the three forward rates in the middle than for $y_t(1)$ and $f_t(5)$. More importantly, both have one single return forecasting factor.

²² We focus on the bias-corrected bootstrap procedure recommended by BH. Small sample bias is a prevailing issue for modeling the yield curve due to its persistence. See Bauer et al. (2012, 2014) for further discussion.

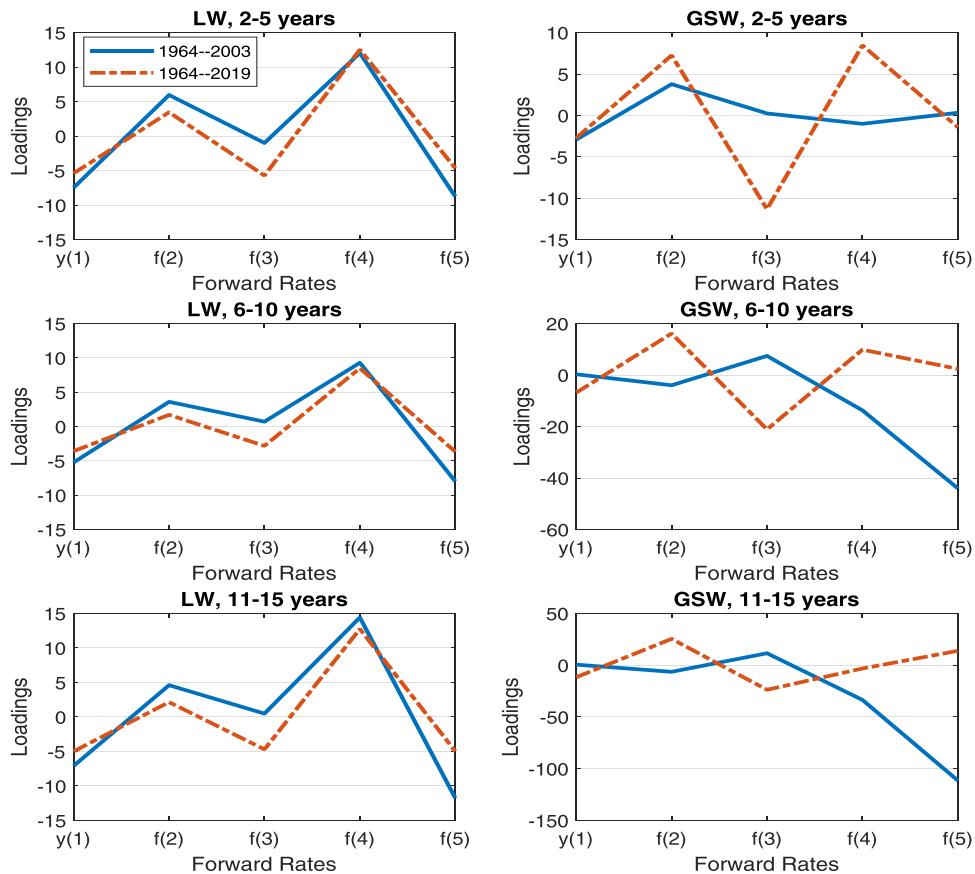


Fig. 5. CP loadings of average excess returns. x-axis: independent variables one- to five-year forward rates $y_t(1)$, $f_t(2), \dots, f_t(5)$. Dependent variables are average excess returns. Bond maturities in the top panels are 2 to 5 years, 6 to 10 years in the middle panels, and 11 to 15 years in the bottom panels. The solid lines are CP's original sample from 1964 to 2003; the dashed lines are the extended sample from 1964 to 2019.

This result is consistent with CP's original conclusion, as well as studies that argue for unspanned factors (e.g., Duffee, 2011).²³ By contrast, the GSW data fail to reject the null hypothesis for CP's original sample from 1964 to 2003, with both p -values at 0.52.

Furthermore, the loadings on PC_4 and PC_5 are stable across the two samples using our data. This result is consistent with the stable loadings illustrated in Figs. 4 and 5. However, with the GSW data, the loading on PC_4 (PC_5) goes from -2.8 (-0.24) to 6.4 (14.7) from CP's original sample to the extended sample. Both the sign and the order of magnitude change, indicating the instability of the loadings.²⁴

Overall, our data support CP's return forecasting factor, and unspanned yield factors. By contrast, the GSW data fail to provide coherent evidence to draw a conclusion in either direction.

5.2. Excess volatility

GK document an excess volatility of long-term bond prices that cannot be explained by the discount-rate variation spanned by shorter-term bond prices.²⁵ Their analysis is based on the GSW data. We repeat their exercises with the GSW data as well as our data.

In an affine term structure model, the short rate is affine in underlying latent factors x_t ,

$$y_t(1) = \delta_0 + \delta_1' x_t, \quad (18)$$

where δ_1 is a column vector.

The factor dynamics follow an AR(1) process under the risk-neutral measure \mathbb{Q} :

$$x_t = \mu^{\mathbb{Q}} + \rho^{\mathbb{Q}} x_{t-1} + \varepsilon_t^{\mathbb{Q}}. \quad (19)$$

As a result, log bond prices are affine in the latent factors,

$$\log P_t(n) = a_n + b_n' x_t, \quad (20)$$

where the loading b_n can be calculated recursively as:

$$b_n = \rho^{\mathbb{Q}'} b_{n-1} - \delta_1, \quad (21)$$

²³ For the Fama and Bliss (1987) data, the null hypothesis is rejected with CP's original sample, but we fail to reject the null with the extended sample using BH bootstrap p -values.

²⁴ Loadings associated with the Fama and Bliss (1987) data also change size and sign, although the change is not as big as with the GSW data.

²⁵ Related work on excess volatility of long-run Treasury bond yields includes Gürkaynak et al. (2005) and Hanson and Stein (2015).

Table 1

CP regression results (predicting 2- to 5-year bond returns). The dependent variable is the one-year excess return $\bar{rx}_{t+1}^{(2 \rightarrow 5)} = \frac{1}{4} \sum_{n=2}^5 rx_{t+1}(n)$. The independent variables are the forward rates: $y_t(1), f_t(2), \dots, f_t(5)$, or the five PCs ($PC_1 - PC_5$). The regression has an intercept. NW t -stats and p -values are computed using Newey-West standard errors with 18 lags. F -stats and the associated p -values report the joint F -test for PC_4 and PC_5 . For both t -stats and F -stats, we also report the corresponding 5% critical values and p -values of the bootstrap distributions of [Bauer and Hamilton \(2018\)](#). p -values are reported in parentheses.

LW

		Forward Rates					PCs					R-square
1964–2003	Loadings	$y(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	PC_1	PC_2	PC_3	PC_4	PC_5	
		-2.100	1.014	1.247	3.632	-3.545	0.164	-1.534	-3.922	3.288	-2.028	0.310
		NW t -stat	-4.013	0.824	0.760	2.708	-3.689	1.977	-4.417	-3.758	2.550	-0.976
	BH 5% c.v.	(0.00)	(0.41)	(0.45)	(0.01)	(0.00)	(0.05)	(0.00)	(0.00)	(0.01)	(0.33)	
		5.117	3.298	2.423	2.133	2.322	4.077	5.838	4.224	2.113	2.290	
	F -stat	(0.17)	(0.70)	(0.55)	(0.01)	(0.00)	(0.40)	(0.16)	(0.09)	(0.02)	(0.41)	
1964–2019	Loadings	BH 5% c.v.										9.051 (0.00)
		F -stat										3.382 (0.00)
		BH 5% c.v.										3.502 (0.00)
	BH 5% c.v.											13.105 (0.00)
												3.502 (0.00)

GSW

		Forward Rates					PCs					R-square
1964–2003	Loadings	$y(1)$	$f(2)$	$f(3)$	$f(4)$	$f(5)$	PC_1	PC_2	PC_3	PC_4	PC_5	
		-2.929	3.774	0.240	-1.016	0.319	0.154	-1.485	3.713	-2.818	-0.240	0.244
		NW t -stat	-4.026	0.613	0.012	-0.039	0.027	1.575	-3.461	2.796	-0.631	-0.007
	BH 5% c.v.	(0.00)	(0.54)	(0.99)	(0.97)	(0.98)	(0.12)	(0.00)	(0.01)	(0.53)	(0.99)	
		4.039	2.358	2.138	2.132	2.156	3.831	5.853	3.595	2.113	2.119	
	F -stat	(0.05)	(0.61)	(0.99)	(0.97)	(0.98)	(0.48)	(0.33)	(0.14)	(0.56)	(0.99)	0.658 (0.52)
1964–2019	Loadings	BH 5% c.v.										3.058 (0.52)
		F -stat										6.623 (0.00)
		BH 5% c.v.										3.042 (0.00)
	BH 5% c.v.											

where $b_1 = -\delta_1$. For derivations, see [Hamilton and Wu \(2012\)](#) and [Creal and Wu \(2015\)](#). The identifying assumptions are ρ^Q is diagonal, $\mu^Q = 0$, and $\delta_1 = [1, 1, 1]'$.

We follow GK's procedure to estimate ρ^Q . First, we regress the log price of the seven-year zero-coupon bond $\log P_t(7)$ on the log prices of one-, three-, and five-year bonds: $\mathcal{P}_t = [\log P_t(1), \log P_t(3), \log P_t(5)]'$, and label the 3×1 slope $\hat{\beta}_7$. Then, we solve the three unknowns in ρ^Q from the three equations for $\hat{\beta}_7$:

$$\hat{\beta}_7 = [b_1, b_3, b_5]^{-1} b_7, \quad (22)$$

where the loadings b_1, b_3, b_5 , and b_7 are functions of ρ^Q through (21). This step of backing out the structural parameters from the reduced-form parameters follows [Hamilton and Wu \(2012\)](#). Let the estimated ρ^Q be $\hat{\rho}^Q$.

For a long-term bond (GK use maturities of 20, 25, and 30 years; we also include 10 and 15 years), we can calculate the amount of price variation explained by \mathcal{P}_t by imposing $\hat{\rho}^Q$:

late the amount of price variation explained by \mathcal{P}_t by imposing $\hat{\rho}^Q$:

$$V^r(n) \equiv \mathbb{V}[\log P_t^r(n)] = \beta_n(\hat{\rho}^Q)' \mathbb{C} \mathbb{O} \mathbb{V}[\mathcal{P}_t] \beta_n(\hat{\rho}^Q), \quad (23)$$

where $\beta_n(\hat{\rho}^Q) = [b_1(\hat{\rho}^Q), b_3(\hat{\rho}^Q), b_5(\hat{\rho}^Q)]^{-1} \beta_n(\hat{\rho}^Q)$. We refer to $V^r(n)$ as the restricted price variation.

Alternatively, we can run an unconstrained OLS regression of the long-term bond price on \mathcal{P}_t , and obtain the unrestricted regression coefficients $\hat{\beta}_n$, and hence the unrestricted price variation $V^u(n) \equiv \mathbb{V}[\log P_t^u(n)] = \hat{\beta}_n' \mathbb{C} \mathbb{O} \mathbb{V}[\mathcal{P}_t] \hat{\beta}_n$. The ratio between the unrestricted and the restricted price variation, which is usually larger than 1, measures the degree of excess volatility:

$$\frac{V^u(n)}{V^r(n)} = \frac{\hat{\beta}_n' \mathbb{C} \mathbb{O} \mathbb{V}[\mathcal{P}_t] \hat{\beta}_n}{\beta_n(\hat{\rho}^Q)' \mathbb{C} \mathbb{O} \mathbb{V}[\mathcal{P}_t] \beta_n(\hat{\rho}^Q)}. \quad (24)$$

We carry out GK's analysis over both their original sample (i.e., 1985–2014) and our extended sample (1985–2019)

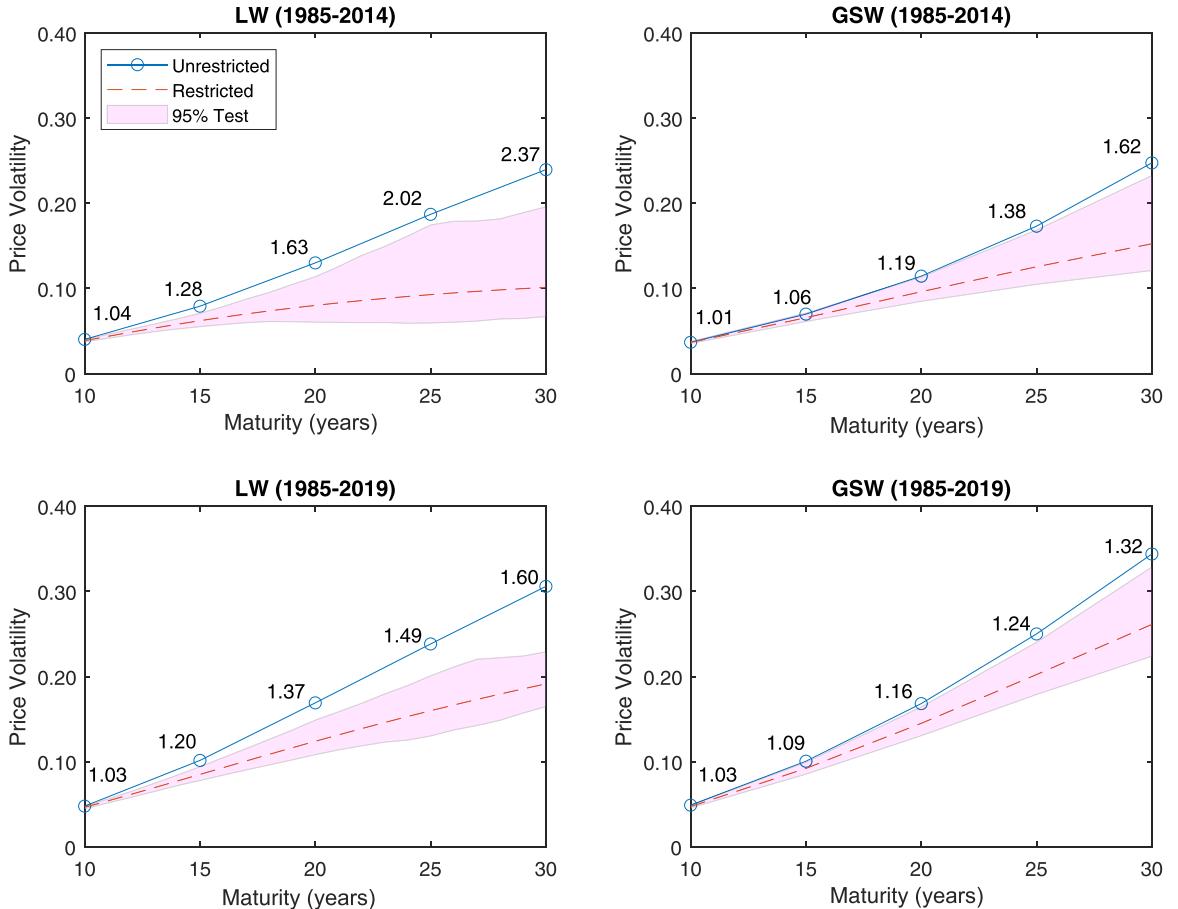


Fig. 6. Excess volatility. We plot the unrestricted variance $V^u(n)$ (solid line) and the restricted variance $V^r(n)$ (dashed line) for the log price of long-term bonds. The shaded area marks the 95% confidence bands. The circles highlight the variance ratios $\frac{V^u(n)}{V^r(n)}$ for bonds with selected maturities.

at the daily frequency.²⁶ We use GK's bootstrap procedure to perform statistical inference, which is constructed under the null hypothesis of no excess volatility; that is, $\frac{V^u(n)}{V^r(n)} = 1$.

We summarize our findings in Fig. 6. We plot the unrestricted price variations $V^u(n)$ in solid lines. The dashed lines show the restricted price variations $V^r(n)$ with a 95% bootstrapped confidence band in the shaded area. The circles mark the maturities of interest, and the numbers above them are variance ratios $\frac{V^u(n)}{V^r(n)}$. Table 2 reports detailed testing results.

In the top-right panel of Fig. 6, we use the GSW data to replicate GK's results for bonds with maturities of 20, 25, and 30 years. Our estimated variance ratios are almost identical to their reported estimates (see their Table II). Long-term bonds display excess volatilities as quantified by a variance ratio of 1.19, 1.38, and 1.62, respectively. Using our data, we find substantially larger estimates than using GSW data, as shown in the top-left panel. In particular, the three variance ratios for long-term bonds are 1.63, 2.02, and 2.37, respectively.

Note GK show a variance ratio larger than 2 is routinely observed for other asset classes, but not for Treasuries. We show this result is driven by the GSW data they use. Using our data therefore reconciles the difference between Treasuries and other assets found in GK in terms of excess price volatility. Moreover, the gap between the solid line and the shaded area is also wider in the top-left panel. Overall, using our data strengthens GK's Treasury results and confirms their main conclusion.

Turning to the extended sample (bottom panels in Fig. 6), variance ratios decline compared with the two panels at the top. This decline can be explained by low and less volatile interest rates caused by the zero lower bound starting from 2009.²⁷ Regardless, the variance ratios esti-

²⁶ More specifically, our daily data are from November 29, 1985, to June 30, 2014, for GK's original sample, and from November 29, 1985 to December 31, 2019, for our extended sample.

²⁷ During the 2009–2015 zero lower bound period, the short end of the yield curve is flat around zero, and does not display much variation. We find a reduction in both the restricted and the unrestricted price variation. However, the unrestricted variation drops more, resulting in a variance ratio of around 0.8 (in particular, 0.82 for 10-year, 0.78 for 15-year, 0.81 for 20-year, 0.77 for 25-year, and 0.87 for 30-year). Post 2015, the market continues to have small price variations under both the restricted and unrestricted model, but the variance ratio is around 1 (in particular, 1.09 for 10-year, 1.04 for 15-year, 0.98 for 20-year, 0.89 for 25-year, and 0.94 for 30-year). Overall, the inclusion of the post-2015 sample for

Table 2

Testing excess volatility. We compute the unrestricted variance $V^u(n)$ and the restricted variance $V^r(n)$ for the log price of long-term bonds. The variance ratio is $\frac{V^u(n)}{V^r(n)}$. p -values and 2.5% and 97.5% critical values are obtained through the bootstrap procedure of [Giglio and Kelly \(2018\)](#), and p -values are reported in parentheses.

Original Sample (1985–2014)

	LW					GSW				
	10yr	15yr	20yr	25yr	30yr	10yr	15yr	20yr	25yr	30yr
Restricted	0.039	0.062	0.080	0.092	0.101	0.037	0.066	0.096	0.126	0.153
2.5% c.v.	0.037	0.054	0.059	0.054	0.062	0.036	0.061	0.085	0.105	0.121
97.5% c.v.	0.040	0.074	0.126	0.189	0.226	0.038	0.071	0.114	0.169	0.233
Unrestricted	0.040	0.079	0.130	0.187	0.239	0.037	0.070	0.115	0.173	0.248
Variance ratio	1.04	1.28	1.63	2.02	2.37	1.01	1.06	1.19	1.38	1.62
	(0.03)	(0.00)	(0.01)	(0.03)	(0.01)	(0.37)	(0.08)	(0.01)	(0.01)	(0.01)
Extended Sample (1985–2019)										
	LW					GSW				
Restricted	0.046	0.085	0.124	0.159	0.191	0.047	0.092	0.145	0.202	0.261
2.5% c.v.	0.044	0.076	0.103	0.120	0.143	0.046	0.085	0.131	0.179	0.224
97.5% c.v.	0.049	0.096	0.151	0.211	0.263	0.049	0.099	0.164	0.240	0.329
Unrestricted	0.048	0.101	0.169	0.238	0.306	0.049	0.101	0.168	0.250	0.344
Variance ratio	1.03	1.20	1.37	1.49	1.60	1.03	1.09	1.16	1.24	1.32
	(0.08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)

mated using our data remain larger than the ones implied by the GSW data.

[Table 2](#) provides test results for the variance ratios. Besides the differences for maturities between 20 and 30 years that we highlight with [Fig. 6](#), we also find that with GK's original sample, a variance ratio of 1 is not rejected (at the 5% level) for the 10-year and 15-year bonds using the GSW data, whereas it is rejected with our data. This observation further corroborates the ubiquity of excess volatility along the maturity spectrum and demonstrates the difference between our data and the GSW data for bonds with intermediate maturities.

Overall, our data provide stronger support for GK's finding of excess volatility for long-term bond prices than the GSW data. The difference stems from the non-parametric method we use as opposed to the parametric approach adopted by GSW to construct the zero-coupon yield curve. In fact, the affine term structure model GK use implies a parametric yield curve with a few parameters. Suppose we use the GK model to construct the zero-coupon yield curve; then, by construction, the variance ratio for all long-term bonds is 1. By contrast, our nonparametric approach allows us to capture unique information that drives the movement of long-term bond prices, leading to a larger estimate of excess volatility.

5.3. Gaussian affine term structure model

Term structure models rely on zero-coupon yield curves to study the economic content of interest rates (e.g., [Ang and Piazzesi, 2003](#); [Joslin et al., 2011](#); [Hamilton and Wu, 2012](#)). In this section, we focus on the workhorse Gaussian affine term-structure model and ask how well it prices raw coupon bonds when we estimate it with our dataset versus GSW's.

the extended sample generates a smaller variance ratio than the 1985–2014 sample. For further details on the zero lower bound, see [Wu and Xia \(2016\)](#).

The one-period risk-free rate $r_t = y_t(1)$ is an affine function of the underlying state variables x_t ,

$$r_t = \delta_0 + \delta'_1 x_t. \quad (25)$$

The state variables follow a first-order vector autoregressive process (VAR(1)) under the physical measure (\mathbb{P}):

$$x_t = \mu + \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma \Sigma'). \quad (26)$$

The price of an n -period bond at time t satisfies:

$$P_t(n) = \mathbb{E}_t[\exp(-m_{t+1})P_{t+1}(n-1)].$$

The log stochastic discount factor is essentially affine as in [Duffee \(2002\)](#):

$$m_{t+1} = r_t + \frac{1}{2}\lambda'_t \lambda_t + \lambda'_t \varepsilon_{t+1}, \quad (27)$$

where the price of risk λ_t is linear in the factors:

$$\lambda_t = \lambda_0 + \lambda_1 x_t.$$

As a result, log bond prices are affine in the factors:

$$\log P_t(n) = a_n + b'_n x_t. \quad (28)$$

The loading a_n, b_n can be calculated recursively as:

$$a_n = a_{n-1} + \mu^\mathbb{Q}' b_{n-1} + \frac{1}{2} b'_{n-1} \Sigma \Sigma' b_{n-1} - \delta_0, \quad (29)$$

$$b_n = \rho^\mathbb{Q}' b_{n-1} - \delta_1, \quad (30)$$

where:

$$\mu^\mathbb{Q} = \mu - \Sigma \lambda_0,$$

$$\rho^\mathbb{Q} = \rho - \Sigma \lambda_1.$$

For derivations, see [Hamilton and Wu \(2012\)](#) and [Creal and Wu \(2015\)](#). The yield of an n -period zero-coupon bond relates to its price as follows:

$$y_t(n) = -\frac{1}{n} \log P_t(n). \quad (31)$$

Following the literature, we have three factors in x_t . The model has a collection of parameters

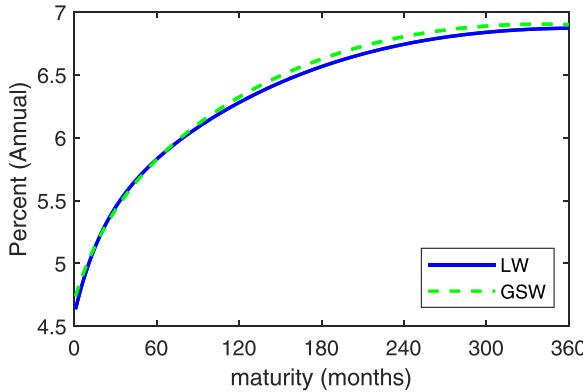


Fig. 7. Unconditional yield curve. The cross-section of the yield curve implied by the term structure model fitted to the LW (in the solid line) and GSW (in the dashed line) zero-coupon yields.

$(\mu, \mu^Q, \rho, \rho^Q, \Sigma, \delta_0, \delta_1)$, and we impose the following identifying assumptions, similar to Joslin et al. (2011) and Hamilton and Wu (2014): (i) $\delta_1 = [1, 1, 1]'$; (ii) $\mu^Q = 0$; (iii) ρ^Q is in real Jordan form with eigenvalues in descending order; and (iv) Σ is lower triangular. The model setup under Q is similar to what we have in Section 5.2, as are the bond prices. The difference is that GK in Section 5.2 focus only on prices, whereas for a term structure model, we jointly model P and Q .

The set of zero-coupon yields we choose to fit are standard. They have maturities of 3 and 6 months, as well as 1, 2, 5, 7, and 10 years. The data are monthly from August 1971 to December 2019. Following the conclusions of Joslin et al. (2011) and Hamilton and Wu (2014), we assume the first three principal components are priced without measurement errors. Fig. 7 shows that both the LW and GSW datasets imply a similar cross-section of the yield curve, with GSW having slightly higher yields between 10 and 30 years.

Fig. 8 demonstrates how well the term structure model fitted to LW (in the solid line) and GSW (in the dashed line) zero-coupon yields price the raw coupon-bearing Treasury securities across different maturity segments. The errors are mean absolute yield error in annualized percentage points. First, we see a U-shape for both lines, which can be explained by the fact that when we estimate the term structure model, we do not fit yields with very short or long maturities. Second, the primary difference between the two lines lies in the short end. The pricing error for securities between 0 and 3 months is 15 bps using our dataset, whereas the corresponding number is 23 bps for the GSW dataset. Third, aggregating across all maturities, the pricing error for LW is 9.1 bps, whereas for GSW, it is 10 bps. The 1 bp difference across maturity segments is likely an understatement of how different these two datasets are, because the term structure model only has three factors. We complement this result with other applications in this section and comparisons of statistical performances in Section 6.

For the very short end, in Fig. 9, we plot the time series to highlight where the differences are. The GSW dataset (the dashed line) has a larger pricing error than the LW

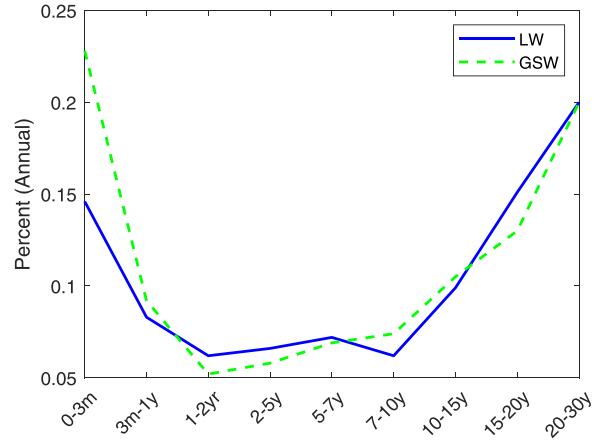


Fig. 8. Pricing errors. Pricing errors calculated as mean absolute yield error in annualized percentage points. Solid line: use LW zero-coupon yields to fit the term structure model, and errors are calculated based on the raw coupon-bearing Treasury securities and the fitted yield curve. Dashed line: use GSW zero-coupon yields to fit the term structure model, and errors are calculated based on the raw coupon-bearing Treasury securities and the fitted yield curve.

dataset (the solid line) throughout the sample. The difference is most prominent prior to the early 1980s, as well as following the Great Recession (2008–2009).

6. Statistical performance of the new yield curve

In Section 5, we discuss the economic implications of our new yield curve, and in this section we turn to its statistical performance. In Section 6.1, we focus on several selected dates to provide some intuition. In Section 6.2, we evaluate the in-sample goodness-of-fit more systematically. In Section 6.3, we present the out-of-sample results. In Section 6.4, we consider alternative specifications for robustness checks.

6.1. Yield curves on selected dates

To gain some insights into the performance of our method, we use YTM computed with the average price to compare our newly constructed yield curve with the raw data as well as the yield curve implied by GSW. For the raw data, the YTM is computed by (6). The model-implied YTM is the solution to (6), except the left-hand side is replaced with the model-implied price \hat{p} , which is defined in (9) for our model.

Fig. 10 illustrates the comparison for four dates: February 1968 (first column), July 2014 (second column), January 1990 (third column), and January 2010 (last column). The top row shows the zero-coupon yield curve; the next two rows are the YTM. Red indicates observations, blue denotes our method, and green is GSW.²⁸

²⁸ We reestimate GSW parameters based on our filtered raw data; see details in A.3. We also use GSW's published parameters (<https://www.federalreserve.gov/data/nominal-yield-curve.htm>) as a robustness check (see C.2) and find the same results.

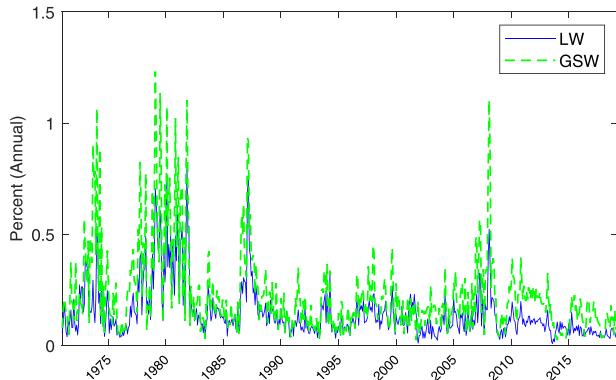


Fig. 9. Time series of pricing errors for maturities under three months. Pricing errors calculated as mean absolute yield error in annualized percentage points. Solid line: use LW zero-coupon yields to fit the term structure model, and errors are calculated based on the raw coupon-bearing Treasury securities (with maturities less than three months) and the fitted yield curve; Dashed line: use GSW zero-coupon yields to fit the term structure model, and errors are calculated based on the raw coupon-bearing Treasury securities (with maturities less than three months) and the fitted yield curve. The sample period is from August 1971 to December 2019.

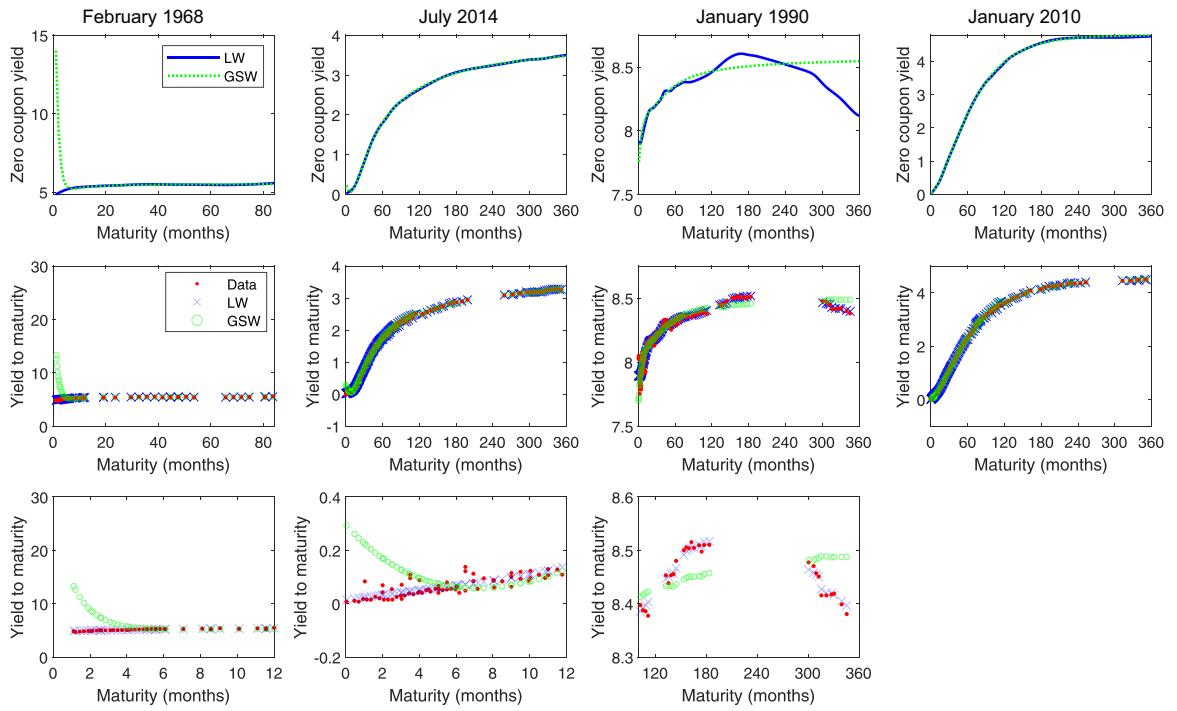


Fig. 10. Yield curves on selected dates. We plot the yield curve estimates, as well as the implied yield to maturities (YTM) for four particular dates: February 1968, July 2014, January 1990, and January 2010. For each date, in the top panels, we plot the two zero-coupon yield curve estimates: one for our method (LW) and one for GSW (GSW). In the middle panels, we plot the YTM of the data (Data) as well as the implied YTM for our method (LW) and GSW (GSW) across all maturities. The bottom panels zoom into a certain maturity range to highlight the difference in YTM.

For February 1968, the main difference between our curve and GSW's is the short end. Our zero-coupon yield approaches around 5% at a maturity of zero, whereas it reaches a level of 15% for GSW's estimate. Importantly, as shown in panel (2,1) and more clearly in panel (3,1), the data do not support GSW's large estimate of the yield at the short end, leading to pricing errors that are in the magnitude of 20%.

The main difference between our model and GSW's in July 2014 is again the short end, which can be better seen in the (3,2) panel. This date lies inside the zero-lower-bound period (from 2009 to 2015). Our curve captures the pattern in the raw data and is consistent with the zero lower bound: the short end converges to zero when maturity approaches zero. By contrast, GSW has a U-shape at the short end, and the difference in YTM between GSW and the raw data is about 0.3%.

Two explanations drive the above results on the short end. First, GSW drop raw data in the short end, including all observations with less than three months to maturity and all Treasury bills. Second, they use a parametric model. Consequently, fitting their parametric model mainly to securities with longer maturities may generate unstable and poorly identified estimates on the short end that are inconsistent with the data.²⁹ By contrast, our non-parametric method allows us to keep raw observations in the short end without sacrificing the fit for longer maturities.

For January 1990, the main difference is for longer maturities; see the (3,3) panel (i.e., the bottom panel of the third column). The raw data display a hump shape, whereas GSW's estimate is monotonically increasing in maturity. Therefore, GSW generate pricing errors that are systematically positive between 300 and 350 months, and negative between 120 and 300 months. Note the 0.1% difference in YTM translates into a 0.5% difference in the zero-coupon yield. By contrast, our estimate fits the raw data well across all maturities.

What drives the performance of GSW is the large gap in the maturity distribution, together with limited observations in the long end, which accounts for a large fraction of our data (see Fig. 1). This feature, combined with parametric methods with a limited degree of freedom, leads to systematic pricing errors for bonds on both sides of the gap. By contrast, our framework allows us to flexibly capture the local variation of the yield curve before and after the gap, resulting in a substantial reduction in pricing errors.

Finally, the last date (January 2010) represents a time when our yield curve estimate agrees with GSW's. Moreover, the implied YTMs from both methods also closely match the data. This example illustrates that our non-parametric method can produce a smooth yield curve and does not overfit.

6.2. In-sample comparisons

In this section, we systematically evaluate the performance of our dataset using the following metrics: root mean-squared pricing error (RMSPE), duration-weighted root-mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted mean absolute pricing error (WMAPE), and mean absolute yield error (MAYE). We also take bid-ask spread into account per (Bliss, 1996) and compile the corresponding MAPE (Bliss), WMAPE (Bliss), and the hit rate (HR (Bliss)). See their definitions in A.4. For the first seven, a smaller error indicates a better model, whereas a larger hit rate is associated with better performance.

Table 3 reports the results of a performance comparison between our method and GSW for nine maturity buckets, together with an overall comparison. Bold text denotes the better performers. We perform this comparison in monthly frequency.

In the top panel, we evaluate our method. The middle panel corresponds to the main GSW specification, where

the zero-coupon yield data are estimated by dropping all securities with maturities less than three months as well as Treasury bills, mimicking GSW's original method and dataset. Our method performs better than GSW across all metrics and maturity buckets. The improvement is substantial, and the reduction in pricing errors across all bonds (last column) ranges between 36% and 63%, with the largest reduction occurring in WMAPE (Bliss). For example, the reduction in MAYE is 55%.

Across maturity buckets, our model performs significantly better at the short end and the longer end. For maturities less than three months, the percentage reduction in the pricing error of our model relative to GSW ranges from 63% to 82%, with WMAPE (Bliss) implying the largest reduction. GSW have large pricing errors in the short end, because they choose not to use the raw data in that segment and subsequently extrapolate the short end from longer-term bonds. For example, they generate large and sometimes extreme pricing errors, with the mean absolute error (in annualized yield to maturity) as high as 7%, whereas our method reduces this number to 0.2%. See more details in B.1.

For maturities above five years, our model again presents a substantial improvement over GSW. The percentage reduction in pricing error ranges from 26% to 34% over the maturity range between 5 years and 10 years, from 48% to 57% between 10 years and 20 years, and 75% to 84% between 20 years and 30 years. We show how the time series compare in B.1.

Note that because GSW do not include securities at the short end when they estimate their model, they do not report the short end of the curve in their dataset. However, researchers often use the parameters provided by GSW to extend the curve at the short end, which is what we have done in the main specification. As a robustness check, in the bottom panel, we estimate the GSW curve including raw data from the short end as an alternative.³⁰ Comparing the two versions, the fitting errors for the short end become smaller with the second version, whereas they are larger over the medium and long maturities, highlighting the tradeoff of GSW's parametric model in fitting both ends of the yield curve. Regardless, our dataset (LW) outperforms both versions of GSW.

6.3. Out-of-sample results

To show the better in-sample performance is not an artificial result of our smooth curve, we next examine its out-of-sample performance. We perform two out-of-sample experiments. The first one is a leave-one-out cross-sectional prediction exercise, and we construct it as follows. On each date, we estimate the yield curve I_t times, where I_t is the number of Treasury securities we observe

²⁹ See Section 5 of GSW for a related discussion on the instability of their estimates.

³⁰ All the results in Section 5 use the main specification of GSW, following what applied researchers typically do: they download the GSW data from their website. Our main specification basically mimics their published data. We show the comparison between the two in Section 6.4.2. Note our main results hold with the alternative version of the GSW data that is estimated including data from the short end; for details, see the Online Appendix.

Table 3

In-sample performance summary. We present results for three models: LW (our model), GSW estimated without data in the short end (which follows their original method), and GSW estimated with data in the short end. For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root-mean-squared pricing error (RMSPE), duration-weighted root-mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the average pricing errors from June 1961 to December 2019.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW										
RMSPE	0.015	0.044	0.073	0.133	0.227	0.391	0.388	0.183	0.116	0.158
WRMSPE	0.012	0.038	0.071	0.125	0.225	0.387	0.387	0.184	0.116	0.051
MAPE	0.012	0.033	0.057	0.101	0.189	0.337	0.343	0.154	0.090	0.084
WMAPE	0.009	0.029	0.056	0.094	0.187	0.333	0.341	0.154	0.091	0.022
MAPE (Bliss)	0.005	0.016	0.028	0.057	0.130	0.257	0.248	0.107	0.057	0.052
WMAPE (Bliss)	0.004	0.014	0.027	0.052	0.128	0.253	0.246	0.108	0.058	0.011
MAYE	0.101	0.063	0.042	0.033	0.038	0.053	0.042	0.012	0.006	0.052
HR (Bliss)	0.424	0.402	0.524	0.403	0.313	0.197	0.290	0.317	0.357	0.412
GSW										
RMSPE	0.041	0.058	0.086	0.173	0.307	0.531	0.625	0.440	0.460	0.246
WRMSPE	0.033	0.053	0.085	0.163	0.305	0.524	0.619	0.442	0.457	0.082
MAPE	0.036	0.046	0.069	0.132	0.265	0.465	0.555	0.399	0.402	0.135
WMAPE	0.028	0.043	0.068	0.124	0.262	0.457	0.548	0.402	0.400	0.041
MAPE (Bliss)	0.028	0.030	0.037	0.085	0.203	0.379	0.451	0.346	0.363	0.101
WMAPE (Bliss)	0.022	0.028	0.037	0.078	0.200	0.371	0.445	0.348	0.361	0.030
MAYE	0.336	0.097	0.051	0.044	0.053	0.071	0.068	0.030	0.028	0.115
HR (Bliss)	0.230	0.341	0.412	0.294	0.184	0.118	0.146	0.114	0.124	0.282
GSW with short-end										
RMSPE	0.016	0.053	0.100	0.192	0.345	0.540	0.684	0.465	0.510	0.262
WRMSPE	0.013	0.045	0.098	0.182	0.341	0.533	0.679	0.467	0.507	0.076
MAPE	0.013	0.040	0.082	0.150	0.299	0.476	0.617	0.428	0.447	0.141
WMAPE	0.009	0.034	0.080	0.142	0.296	0.469	0.611	0.430	0.444	0.028
MAPE (Bliss)	0.006	0.022	0.048	0.101	0.236	0.392	0.511	0.373	0.407	0.106
WMAPE (Bliss)	0.004	0.018	0.047	0.094	0.233	0.385	0.505	0.375	0.404	0.017
MAYE	0.108	0.074	0.060	0.050	0.059	0.073	0.074	0.033	0.031	0.066
HR (Bliss)	0.410	0.349	0.344	0.249	0.165	0.120	0.128	0.091	0.094	0.296

at t . Each time, we leave out a bond i and use the remaining bonds to compute the model-implied price for bond i and hence its out-of-sample pricing error using the same eight metrics as in Section 6.2. We repeat this exercise over t to calculate the average pricing error. Given the computational burden of the out-of-sample exercise, we use a quarterly sample from 1961:Q2 to 2019:Q4.³¹

We report in Table 4 the leave-one-out out-of-sample comparison between our approach and GSW's. The out-of-sample results are similar to the results for the in-sample comparison in Table 3: our method produces smaller pricing errors and a higher hit rate across all maturity buckets, metrics, and two versions of GSW.

Focusing on the main specification of GSW in the middle panel of Table 4, the average reduction in MAYE is 49%. Fig. 11 shows the time series of the out-of-sample pricing error for maturities less than three months. Our model implies smaller out-of-sample pricing errors across all time periods, suggesting the inability of the main GSW specification to capture the very short end of the yield

curve. For other maturities, although the relative performance changes at times, our model still compares favorably to GSW because GSW exhibits more spikes in pricing error (see B.2).

Turning to the results in the bottom panel of Table 4 (i.e., GSW estimated with the short end included), similar to Table 3, the results demonstrate an improved performance over the short end compared with the results in the middle panel. However, such an improvement comes at the cost of a larger pricing error over the medium and long maturities. For example, between one year and five years, where the original GSW approach puts a high weight on when minimizing pricing errors, the out-of-sample pricing error in the bottom panel increases by around 25% compared to the original GSW approach (middle panel). Despite the differences between the two versions of GSW, our dataset still outperforms GSW.

The second out-of-sample comparison is based on time series prediction. Specifically, we use the estimated curve at the end of month t to predict the curve at the beginning of month $t+1$. We report the results in Table 5. Similar to the in-sample comparison in Table 3 and the leave-one-out comparison in Table 4, our dataset also performs better than both versions of GSW in terms of the time series prediction across all maturities and evaluation metrics.

³¹ For each of the four model specifications, the out-of-sample exercise takes about two weeks for a computer with dual Intel Xeon Gold 6136 CPU and 208 GB memory using 24 workers in MATLAB parallel computing.

Table 4

Leave-one-out comparison. We present results for three models: LW (our model), GSW estimated without data in the short end (which follows their original method), and GSW estimated with data in the short end. For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root mean-squared pricing error (RMSPE), duration-weighted root mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the averaged pricing errors over the full sample from June 1961 to December 2019 at the quarterly frequency.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW										
RMSPE	0.015	0.045	0.085	0.158	0.281	0.486	0.549	0.238	0.163	0.199
WRMSPE	0.013	0.040	0.083	0.148	0.279	0.480	0.548	0.239	0.163	0.062
MAPE	0.012	0.034	0.066	0.118	0.235	0.415	0.492	0.198	0.126	0.100
WMAPE	0.010	0.030	0.065	0.110	0.232	0.410	0.491	0.199	0.126	0.024
MAPE (Bliss)	0.005	0.017	0.036	0.072	0.173	0.333	0.388	0.150	0.091	0.067
WMAPE (Bliss)	0.004	0.015	0.035	0.066	0.170	0.327	0.387	0.151	0.091	0.013
MAYE	0.103	0.066	0.049	0.039	0.047	0.065	0.061	0.016	0.009	0.056
HR (Bliss)	0.438	0.389	0.492	0.360	0.246	0.162	0.229	0.268	0.270	0.386
GSW										
RMSPE	0.037	0.050	0.091	0.188	0.330	0.564	0.608	0.475	0.467	0.283
WRMSPE	0.032	0.046	0.089	0.175	0.331	0.546	0.604	0.462	0.465	0.094
MAPE	0.034	0.040	0.072	0.141	0.291	0.485	0.521	0.424	0.403	0.163
WMAPE	0.027	0.037	0.071	0.133	0.282	0.473	0.538	0.432	0.401	0.044
MAPE (Bliss)	0.031	0.015	0.039	0.093	0.227	0.400	0.416	0.371	0.365	0.121
WMAPE (Bliss)	0.023	0.013	0.039	0.085	0.219	0.387	0.411	0.372	0.362	0.031
MAYE	0.317	0.070	0.051	0.047	0.055	0.074	0.065	0.032	0.027	0.112
HR (Bliss)	0.129	0.541	0.388	0.282	0.178	0.092	0.153	0.101	0.119	0.289
GSW with short-end										
RMSPE	0.016	0.048	0.111	0.207	0.361	0.577	0.655	0.501	0.527	0.303
WRMSPE	0.014	0.045	0.108	0.192	0.372	0.554	0.660	0.487	0.514	0.091
MAPE	0.013	0.037	0.085	0.163	0.320	0.494	0.575	0.443	0.467	0.170
WMAPE	0.010	0.035	0.082	0.154	0.316	0.486	0.599	0.452	0.449	0.034
MAPE (Bliss)	0.006	0.014	0.051	0.109	0.257	0.415	0.458	0.392	0.403	0.127
WMAPE (Bliss)	0.004	0.012	0.049	0.105	0.233	0.396	0.463	0.397	0.409	0.021
MAYE	0.110	0.068	0.063	0.059	0.061	0.079	0.072	0.036	0.031	0.073
HR (Bliss)	0.411	0.557	0.327	0.221	0.147	0.090	0.119	0.085	0.088	0.295

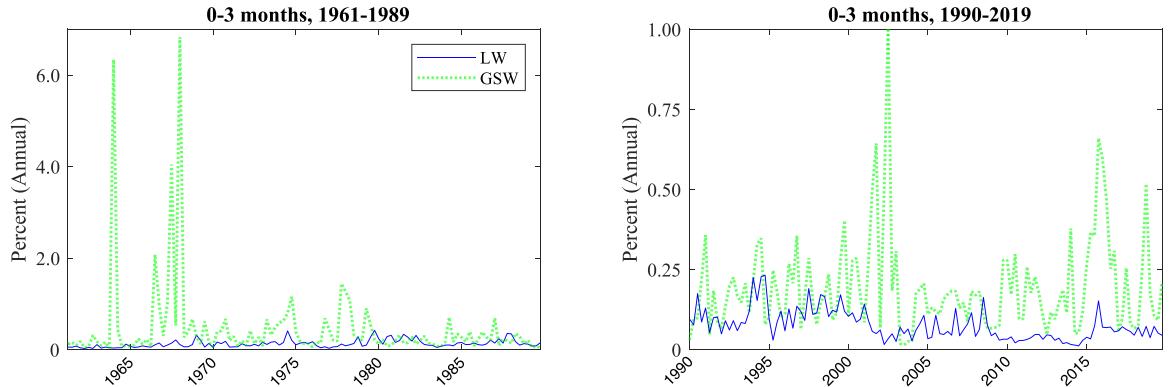


Fig. 11. Time series of mean absolute error in YTM: The short end, leave-one-out. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over the entire sample period (i.e., 1961–2019). We calculate the mean absolute pricing error across maturities between zero and three months. We report out-of-sample pricing errors at the quarterly frequency.

6.4. Robustness

We further explore several alternative specifications of our model in Section 6.4.1 and of GSW in Section 6.4.2 for robustness checks.

6.4.1. Alternative specifications of our model

First, one may be concerned with data quality over a particular segment of the maturity spectrum. For example, one may believe that Treasuries with maturities below one year should be priced differently than other maturities

Table 5

Time-series prediction comparison. We present results for three models: LW (our model), GSW estimated without data in the short end (which follows their original method), and GSW estimated with data in the short end. We perform an out-of-sample prediction exercise where the yield curve estimated at the end of month t is used to price bonds at the beginning of month $t+1$. For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root-mean-squared pricing error (RMSPE), duration-weighted root-mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the averaged pricing errors over the full sample from June 1961 to December 2019 at the monthly frequency.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW										
RMSPE	0.017	0.053	0.111	0.230	0.389	0.604	0.743	0.726	0.755	0.320
WRMSPE	0.013	0.046	0.108	0.217	0.386	0.600	0.740	0.727	0.755	0.088
MAPE	0.013	0.040	0.094	0.193	0.348	0.545	0.699	0.699	0.739	0.192
WMAPE	0.009	0.035	0.091	0.181	0.345	0.542	0.697	0.700	0.739	0.031
MAPE (Bliss)	0.006	0.023	0.059	0.144	0.285	0.463	0.596	0.644	0.698	0.156
WMAPE (Bliss)	0.004	0.020	0.057	0.133	0.282	0.459	0.593	0.645	0.698	0.021
MAYE	0.109	0.076	0.066	0.062	0.069	0.083	0.077	0.053	0.047	0.075
HR (Bliss)	0.425	0.359	0.333	0.215	0.149	0.092	0.108	0.066	0.042	0.276
GSW										
RMSPE	0.040	0.067	0.119	0.255	0.442	0.687	0.917	0.837	0.925	0.379
WRMSPE	0.031	0.060	0.116	0.240	0.439	0.681	0.909	0.838	0.925	0.110
MAPE	0.035	0.054	0.101	0.210	0.399	0.624	0.851	0.804	0.868	0.222
WMAPE	0.025	0.049	0.098	0.197	0.396	0.618	0.844	0.806	0.868	0.047
MAPE (Bliss)	0.027	0.036	0.065	0.160	0.335	0.537	0.744	0.749	0.827	0.185
WMAPE (Bliss)	0.019	0.033	0.062	0.148	0.332	0.531	0.736	0.751	0.827	0.036
MAYE	0.333	0.108	0.070	0.068	0.078	0.094	0.097	0.061	0.056	0.135
HR (Bliss)	0.242	0.292	0.290	0.198	0.122	0.071	0.078	0.061	0.036	0.213
GSW with short-end										
RMSPE	0.019	0.061	0.130	0.268	0.473	0.691	0.961	0.851	0.953	0.390
WRMSPE	0.014	0.052	0.127	0.253	0.469	0.685	0.954	0.853	0.952	0.104
MAPE	0.014	0.047	0.111	0.222	0.427	0.629	0.895	0.819	0.891	0.223
WMAPE	0.008	0.040	0.108	0.208	0.424	0.624	0.888	0.821	0.891	0.035
MAPE (Bliss)	0.008	0.029	0.074	0.171	0.362	0.543	0.786	0.764	0.850	0.186
WMAPE (Bliss)	0.005	0.024	0.071	0.159	0.359	0.538	0.780	0.765	0.849	0.025
MAYE	0.118	0.086	0.077	0.072	0.084	0.095	0.101	0.062	0.058	0.085
HR (Bliss)	0.407	0.315	0.262	0.185	0.108	0.078	0.078	0.062	0.031	0.244

for institutional reasons. While we believe keeping information from the short end disciplines the overall behavior of the yield curve, we can adapt our approach by altering the weights we put on different maturities. We therefore experiment with a setup where we put less weight on short maturities. In particular, we alter our objective function in (7) with a weight of $\frac{1}{\gamma D_i^2}$. We set $\gamma = 2$ for maturities less than one year to downweight information from the short end, and $\gamma = 1$ for all other maturities. We report the performance summary in Tables C.1 of C.1. Although we use a different weighting scheme to estimate the yield curve, our performance metrics (especially those that use bond durations) are kept the same as before. Compared with our baseline results in Table 3, our results in Table C.1 show that putting less weights on the short end leads to slightly larger pricing errors on the short end and somewhat smaller pricing errors on the long end (e.g., for maturities between seven years and ten years). Regardless, the zero-coupon yields compiled by our method with different weights still outperform the zero-coupon yields of GSW across maturities.

Second, in our baseline results, we use an outlier-detection algorithm that uses the 3.0 inter-quartile range

(IQR) rule; that is, we discard observations that are either 3.0 IQR above the 75th percentile or 3.0 IQR below the 25th percentile for a local maturity bucket (see more details in A.2). This rule is less aggressive (in terms of discarding outliers) than usual because our intention is to keep as many observations as possible. One may wonder whether alternative IQR cutoffs would materially change our results. We thus implement an alternative of 2.0 IQR, which is more aggressive than our baseline specification, by discarding more outliers. The results are in Tables C.2 of C.1. Dropping more outliers makes the data less noisy, leading to a smaller MAYE; for example, the MAYE is 0.049% across maturity buckets, whereas it is 0.052% in Table 3. However, this does not change the main result in Table 3 that our dataset yields better performance than GSW's across all maturity buckets and all performance metrics.

Third, we also compare alternative values for N_0 in an out-of-sample exercise to search for the optimal bandwidth (see details in Table C.3 of C.1). Different values of N_0 do not change the relative performance between our approach and GSW's. Across different values, $N_0 = 8$ generates the smallest pricing error in terms of RMSPE, and leads to a smoother yield curve than smaller N_0 values. We

therefore consider $N_0 = 8$ the optimal bandwidth parameter.

6.4.2. Alternative specifications of GSW

To show our comparison with GSW in [Table 3](#) is not an artifact of the specifics of how we estimate the GSW curve, we implement an alternative by using the published GSW parameter values from the Federal Reserve's website: <https://www.federalreserve.gov/data/nominal-yield-curve.htm>. The results are in [Tables C.4](#) of [C.2](#). The first two panels are the benchmark LW and GSW, which are the same as in [Table 3](#). The alternative GSW specification with their published parameters yields larger pricing errors and lower hit rates compared to the benchmark GSW specification. Therefore, the comparison favors our data even more.

7. Conclusion

The zero-coupon yield curve provides important information about financial markets and the macroeconomy, and is widely used by researchers and practitioners. In this paper, we develop a new dataset using a non-parametric kernel-smoothing method with a novel adaptive bandwidth specifically designed to fit the US Treasury yield curve. We show our yield curve estimates provide a more accurate description of the raw data than the leading alternative, ([Gürkaynak et al., 2007](#)). More importantly, we show our reconstructed yield curve leads to different conclusions than [Gürkaynak et al. \(2007\)](#) in two influential studies: [Cochrane and Piazzesi \(2005\)](#) and [Giglio and Kelly \(2018\)](#).

Appendix A. Details on implementation

A.1. Kernel-smoothing method: Derivation and estimation

The first-order conditions for the minimization problem in [\(7\)](#) are:

$$\sum_{i=1}^I \sum_{j=1}^{J_i} \Phi_j^i(n; y, y') \cdot 1/D_i^2 = 0, \quad (\text{A.1})$$

$$\sum_{i=1}^I \sum_{j=1}^{J_i} \Phi_j^i(n; y, y') (n - v_j^i) \cdot 1/D_i^2 = 0, \quad (\text{A.2})$$

where $\Phi_j^i(n; y, y')$ is given by:

$$\begin{aligned} & \Phi_j^i(n; y, y') \\ &= \left(K_{h(v_j^i)}(n - v_j^i) c_j^i v_j^i d_j^i(n) \right) \\ & \times \left(p^i - c_j^i d_j^i(n) - \sum_{k=1, k \neq j}^I \left(\int K_{h(v_k^i)}(n - v_k^i) c_k^i d_k^i(n) dn \right) \right), \end{aligned} \quad (\text{A.3})$$

$$d_k^i(n) = \exp \left[- \left(y(n) + (v_k^i - n)y'(n) \right) v_k^i \right]. \quad (\text{A.4})$$

Note [Eq. \(A.3\)](#) (and therefore [Eqs. \(A.1\)](#) and [\(A.2\)](#)) contains integrals. Although, in principle, solving [Eqs. \(A.1\)](#) and [\(A.2\)](#) numerically is possible,³² we follow [Jeffrey et al. \(2006\)](#) and approximate the integrals with interpolations that are functions of $y(\cdot)$ and $y'(\cdot)$.³³

In particular, suppose the support of $y(\cdot)$ and $y'(\cdot)$ is $\mathcal{N} = \{1, 2, \dots, 360\}$. We approximate the integrals in [Eq. \(A.3\)](#) as:

$$\approx \frac{\int K_{h(v_k^i)}(n - v_k^i) d_k^i(n) dn}{\sum_{n=1}^{360} K_{h(v_k^i)}(n - v_k^i)}. \quad (\text{A.5})$$

On the other hand, viewing $d_k^i(n)$ as the discount rate at v_k^i approximated by the yield curve at n , $\int K_{h(v_k^i)}(n - v_k^i) d_k^i(n) dn$ can be interpreted as the kernel-smoothed discount rate at v_k^i . Letting the corresponding zero-coupon yield be $\hat{y}(v_k^i)$, which is defined through $\int K_{h(v_k^i)}(n - v_k^i) d_k^i(n) dn = \exp[-v_k^i \times \hat{y}(v_k^i)]$, we obtain $\hat{y}(v_k^i)$ as³⁴:

$$\begin{aligned} & \hat{y}(v_k^i) \\ &= -\frac{1}{v_k^i} \log \left(\frac{\sum_{n=1}^{360} K_{h(v_k^i)}(n - v_k^i) \exp \left[-\left(y(n) + (v_k^i - n)y'(n) \right) v_k^i \right]}{\sum_{n=1}^{360} K_{h(v_k^i)}(n - v_k^i)} \right). \end{aligned} \quad (\text{A.6})$$

Replacing v_k^i by an arbitrary maturity v in [\(A.6\)](#), we arrive at the formula that we use to interpolate the yield curve at any maturity v .

In sum, we seek to solve [Eqs. \(A.1\)](#) and [\(A.2\)](#) with respect to $y(n)$ and $y'(n)$ for $n \in \mathcal{N} = \{1, 2, \dots, 360\}$, where $\Phi_j^i(n; y, y')$ is given by [Eqs. \(A.3\)](#) and [\(A.4\)](#), but with the integrals in [Eq. \(A.3\)](#) replaced by [Eqs. \(A.5\)](#) and [\(A.6\)](#). In essence, we are solving a system of non-linear equations. By construction, all of these equations involve functions that are infinitely differentiable. We provide closed-form gradients for these equations,³⁵ which allows us to solve these non-linear equations efficiently.

A.2. Outlier detection

Our outlier-detection algorithm follows several steps.

³² See [Linton et al. \(2001\)](#) for the iterative algorithms they propose to solve a system of equations that are similar to [Eqs. \(A.1\)](#) and [\(A.2\)](#).

³³ We implemented both [Linton et al. \(2001\)](#) (in particular, the log-linear specification) and [Jeffrey et al. \(2006\)](#) for our model. Our experience is that [Jeffrey et al. \(2006\)](#) indeed offer a more stable and computationally efficient solution than [Linton et al. \(2001\)](#).

³⁴ The above interpolation can be interpreted as the solution to an optimization problem that is similar to [\(7\)](#) for a pure discount bond with a maturity of v_k^i . More specifically, for a given estimated $\tilde{y}(\cdot)$ and $\tilde{y}'(\cdot)$, the solution to the minimization problem $\min_{y(v_k^i)} \int \left(\exp \left[-y(v_k^i) \times v_k^i \right] - \exp \left[-\left(\tilde{y}(n) + (v_k^i - n)\tilde{y}'(n) \right) n \right] \right)^2 K_{h(v_k^i)}(n - v_k^i) dn$ is given by $y(v_k^i) = -\frac{1}{v_k^i} \log \left(\int \left(\exp \left[-\left(\tilde{y}(n) + (v_k^i - n)\tilde{y}'(n) \right) n \right] \right) K_{h(v_k^i)}(n - v_k^i) dn \right)$. Because we only have solutions for $\tilde{y}(\cdot)$ and $\tilde{y}'(\cdot)$ over $\mathcal{N} = \{1, 2, \dots, 360\}$, the interpolated version of this solution is given by [Eq. \(A.6\)](#).

³⁵ This is another benefit of replacing the integrals in [Eq. \(A.3\)](#) with interpolated yields as in [Eq. \(A.6\)](#).

First, we drop observations whose yield to maturity (YTM) is higher than 30% (annualized). In the data, sometimes bond price appears to be too low (equivalently, YTM appears to be too high). Across time, bond prices in general reach their lowest during the early 1980s recession, approaching a level of 20% in YTM. We therefore set $30\% = 1.5 \times 20\%$ as a lenient threshold in YTM to drop low price observations. Note potential outliers that have a high YTM but still below 30%, which are not dropped after this step, are likely to be dropped after the following steps.

Next, suppose the current day is t . We obtain our non-parametric yield curve estimate from day $t - 1$. Based on this estimate, we calculate the implied YTM for all bond observations for day t (denoted as $\widehat{YTM}_{i,t-1}$ for bond i). Assuming yield curves are internally consistent across days, and therefore display relatively small day-to-day variations, we take $\widehat{YTM}_{i,t-1}$ as the benchmark YTM and evaluate the distance between $YTM_{i,t}$ (i.e., current YTM for bond i) and $\widehat{YTM}_{i,t-1}$. A bond with a large distance is a suspect outlier.

To take into account the difference in data quality (i.e., noise level) in different maturity segments, we group bonds into several maturity ranges. For each maturity range and for each bond within, we calculate $Dist_{i,t} \equiv |YTM_{i,t} - \widehat{YTM}_{i,t-1}|$. We use $\pm Dist_{i,t}$ for all bonds within the maturity range, and calculate the interquartile range (IQR, i.e., the 75th percentile minus the 25th percentile). An outlier is detected if its current-day YTM (i.e., $YTM_{i,t}$) is either below $\widehat{YTM}_{i,t-1} - 3 \times IQR$ or above $\widehat{YTM}_{i,t-1} + 3 \times IQR$. We choose three [rather than 1.5 as usual for outlier detection; see, for example, Tukey (1977)] to be conservative in excluding bond observations. We also choose three maturity segments: bonds with a maturity of less than one year, between one year and five years, and above five years. Our 3.0-IQR rule applied to segment-specific data allows us to keep as much data as possible. Bond observations that are identified as outliers often have a large discrepancy (in terms of YTM) from other observations that have similar maturities.

Overall, our outlier-detection method allows us to drop, on average, three observations for each day.

A.3. Our implementation of GSW

We obtain GSW's parameters from the Federal Reserve Board's webpage, and use them as starting values and re-estimate their model based on our raw bond data. Besides applying our filters described in Section 4.1, we also drop securities with less than three months to maturity and all Treasury bills, following GSW. In addition, we follow GSW by using the Nelson-Siegel four-parameter specification for the period before 1980 and GSW's six-parameter specification for the post-1980 period.

For most months, the re-estimated GSW curve is very similar to their original curve computed using their pub-

lished parameters. This finding confirms the similarity in the underlying raw data we use. For a few months, the two versions have a substantial difference in the short end, where observations are omitted in estimation following GSW. This instability is consistent with what GSW find; see Section 5 of their paper. Given the parameter instability of GSW, we compare our method with both the re-estimated GSW and their reported parameters.

A.4. Model comparison metrics

Let the actual bond price and the model-implied bond price be p_i and \widehat{p}_i for $i = 1, 2, \dots, I$. We first define two measures of pricing error that are directly related to our objective function. The first is the root mean-squared pricing error (RMSPE) that calculates the square root of the mean-squared distance between p_i and \widehat{p}_i , that is,

$\sqrt{\frac{1}{I} \sum_{i=1}^I (p_i - \widehat{p}_i)^2}$. The second is the duration-weighted root mean-squared pricing error (WRMSPE) defined as

$\sqrt{\frac{1}{I} \sum_{i=1}^I w_i^2 (p_i - \widehat{p}_i)^2}$, where $w_i = \frac{D_i^{-1}}{\sum_{i=1}^I D_i^{-1}}$ is the weight for bond i . Note WRMSPE is equivalent to our objective function that also weights pricing errors by bond durations.

We next define two metrics related to absolute pricing errors. They are the mean absolute pricing error, $MAPE = \frac{1}{I} \sum_{i=1}^I |p_i - \widehat{p}_i|$, and the duration-weighted mean absolute pricing error, $WMAPE = \sum_{i=1}^I w_i |p_i - \widehat{p}_i|$.

Bliss (1996) argues the bid-ask spread needs to be taken into account when calculating the pricing error. We follow Bliss (1996) to define the bid-ask-spread-adjusted pricing error as:

$$\varepsilon_i = \begin{cases} \widehat{p}_i - p_i^a & \text{if } \widehat{p}_i > p_i^a, \\ p_i^b - \widehat{p}_i & \text{if } \widehat{p}_i < p_i^b, \\ 0 & \text{otherwise,} \end{cases}$$

where p_i^a and p_i^b are the ask and bid quotes, respectively, for the bond. The corresponding mean absolute pricing error [denoted as MAPE (Bliss)] and duration-weighted absolute pricing error [denoted as WMAPE (Bliss)] are defined as $\frac{1}{I} \sum_{i=1}^I \varepsilon_i$ and $\sum_{i=1}^I w_i \varepsilon_i$.

Next, rather than calculating the error between the actual and the fitted price, we define the mean absolute yield error (MAYE) as the average absolute error between the observed and the fitted YTM.

Lastly, we follow Bliss (1996) to define the hit rate (HR (Bliss)) as $\frac{1}{I} \sum_{i=1}^I \mathbb{1}_{\{p_i^b \leq \widehat{p}_i \leq p_i^a\}}$, where $\mathbb{1}_{\{p_i^b \leq \widehat{p}_i \leq p_i^a\}}$ is the indicator function that equals 1 if \widehat{p}_i falls within $[p_i^b, p_i^a]$. The hit rate calculates the number of times the fitted price falls within the bid-ask spread.

Appendix B. Time series evidence

B.1. In sample

In this section, we examine the time series of pricing errors to provide more insights into the performance of our method.

The short term. Fig. B.1 shows our model performs consistently better than GSW across different time periods for maturities less than one year. The top panel plots maturities less than three months, and the bottom panel plots maturities between three months and one year. We split the full sample into the 1961–1989 (left panels) and the 1990–2019 (right panels) sub-samples given the general decline in pricing error over time.³⁶

For maturities less than three months (top panels), we observe that GSW occasionally generate large pricing errors at around 7 annualized percentage points. The left column of Fig. 10 illustrates one such example in February 1968. We see a general decline in pricing error over the post-1990 period. However, GSW's pricing error can still reach 1%. Our method is able to reduce these pricing errors significantly.

For maturities between three months and one year (bottom panels), our model continues to outperform GSW. In particular, our model does better than GSW for the 1961–1975 period and the more recent post-2009 period. The post-2009 period is associated with the zero lower bound and a subsequent low interest rate environment. As in the second column of Fig. 10, we illustrate that our method fits the short end of the yield curve better for this period.³⁷

The large pricing errors of GSW at the short end come from the fact that they exclude all securities with less than three months to maturity as well as all Treasury bills. Consequently, GSW extrapolate the short end of the yield curve from securities with longer maturities, which leads to imprecise and sometimes extreme estimates of the short end of the yield curve.

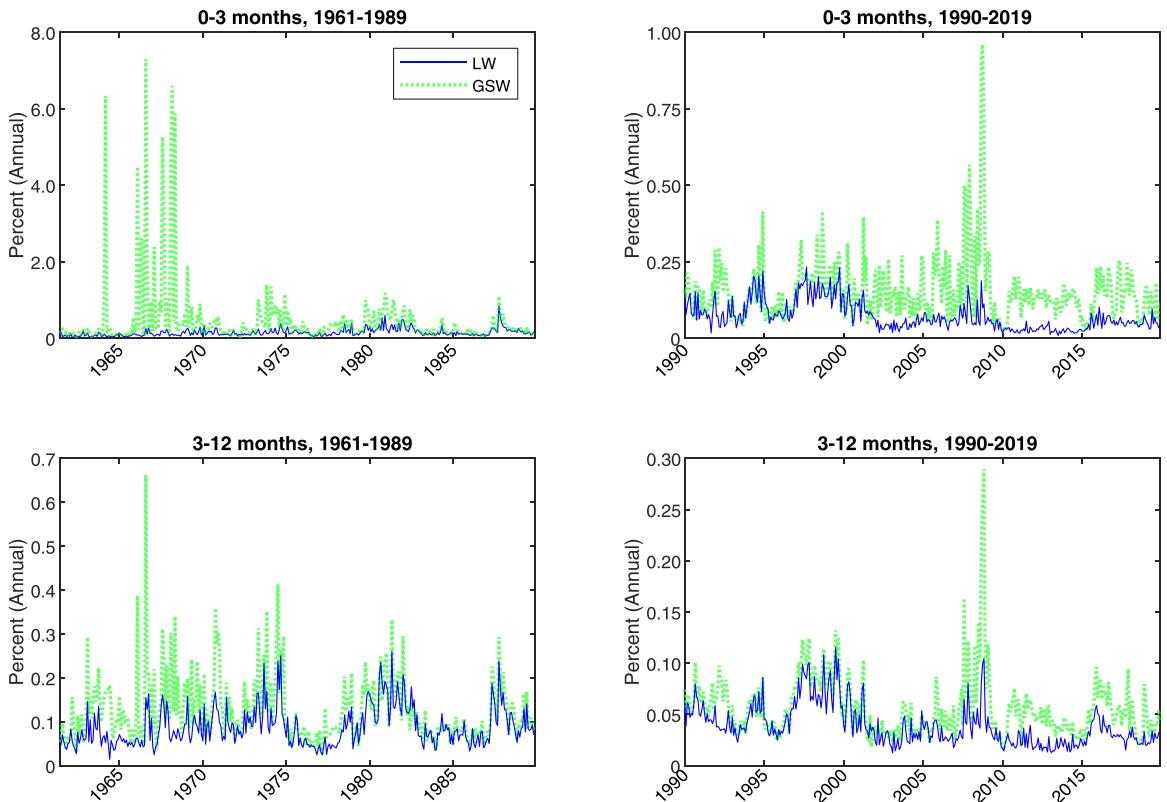


Fig. B.1. Time series of mean absolute error in YTM: The short end. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over the entire sample period (i.e., 1961–2019). The top panel shows bonds with maturities less than three months; the bottom panel shows bonds with maturities between three months and one year.

³⁶ To ensure that reestimating GSW's parameters based on our data does not cause the large pricing errors for their model, we plot the minimum pricing error between their original and our reestimated versions in Fig. B.1.

³⁷ The superior performance of the short end of our yield curve data makes them particularly useful for studies that try to disentangle short-rate expectations and risk premiums in driving bond returns [see Cieslak (2018) for a recent application].

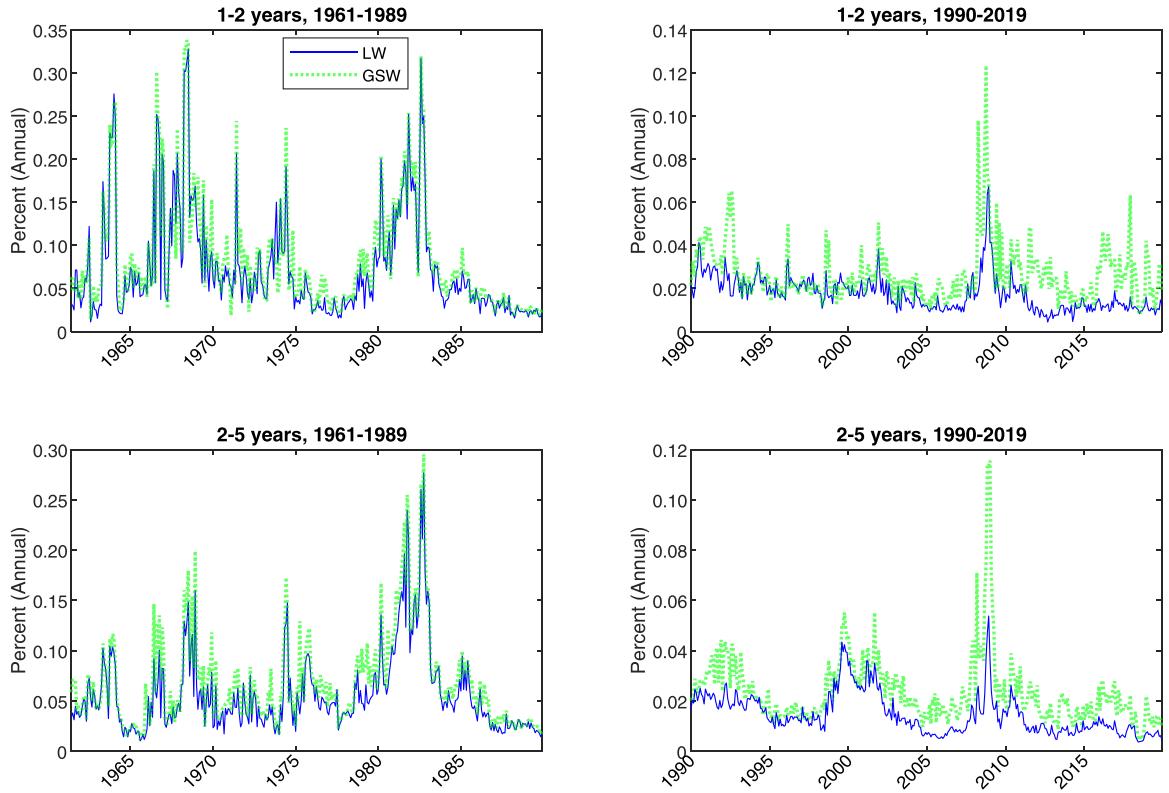


Fig. B.2. Time series of mean absolute error in YTM: The medium end. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over 1961–1989 (left panels) and 1990–2019 (right panels). We group bonds into two maturity buckets: 1–2 years (top row) and 2–5 years (bottom row).

Moreover, the issue of the short end of the yield curve of GSW is unlikely to be solved by simply including securities with short maturities in their estimation. The challenge is that the parametric model used in GSW has a limited degree of freedom and cannot simultaneously capture short-term, medium-term, and long-term yields.

By contrast, our non-parametric framework with adaptive bandwidth presents a natural solution to this challenge, because it adjusts the amount of local information used to construct the yield curve.

The medium and long term

For maturities between one and five years, our model performs similarly to GSW, with the exception of the post-2009 sample (see Fig. B.2). During the low interest rate period after 2009, our model significantly outperforms GSW's. The similarity in performance before that period is consistent with the observation that abundant data are available over this maturity range, causing parametric models such as GSW's to use most of its degree of freedom to fit this part of the data.

For maturities above five years, we again see substantial improvement of our model over GSW (see Fig. B.3). Between five years and ten years, we start to see the improvement of our method. For example, between 2000 and 2006, we are able to reduce the MAYE from 0.08% in GSW to around 0.02%.

The pricing errors of GSW for maturities longer than 10 years contain large spikes. For example, between 1986 and 1990, GSW's pricing error for maturities between 20 years and 30 years reaches 0.3%. By contrast, the pricing error from our method always stays under 0.05%. Moreover, our improvement applies not only to the pre-1990 sample for which a limited number of long-term securities are outstanding, but also to the post-1990 sample, including the most recent sample when abundant data on the long end are available.

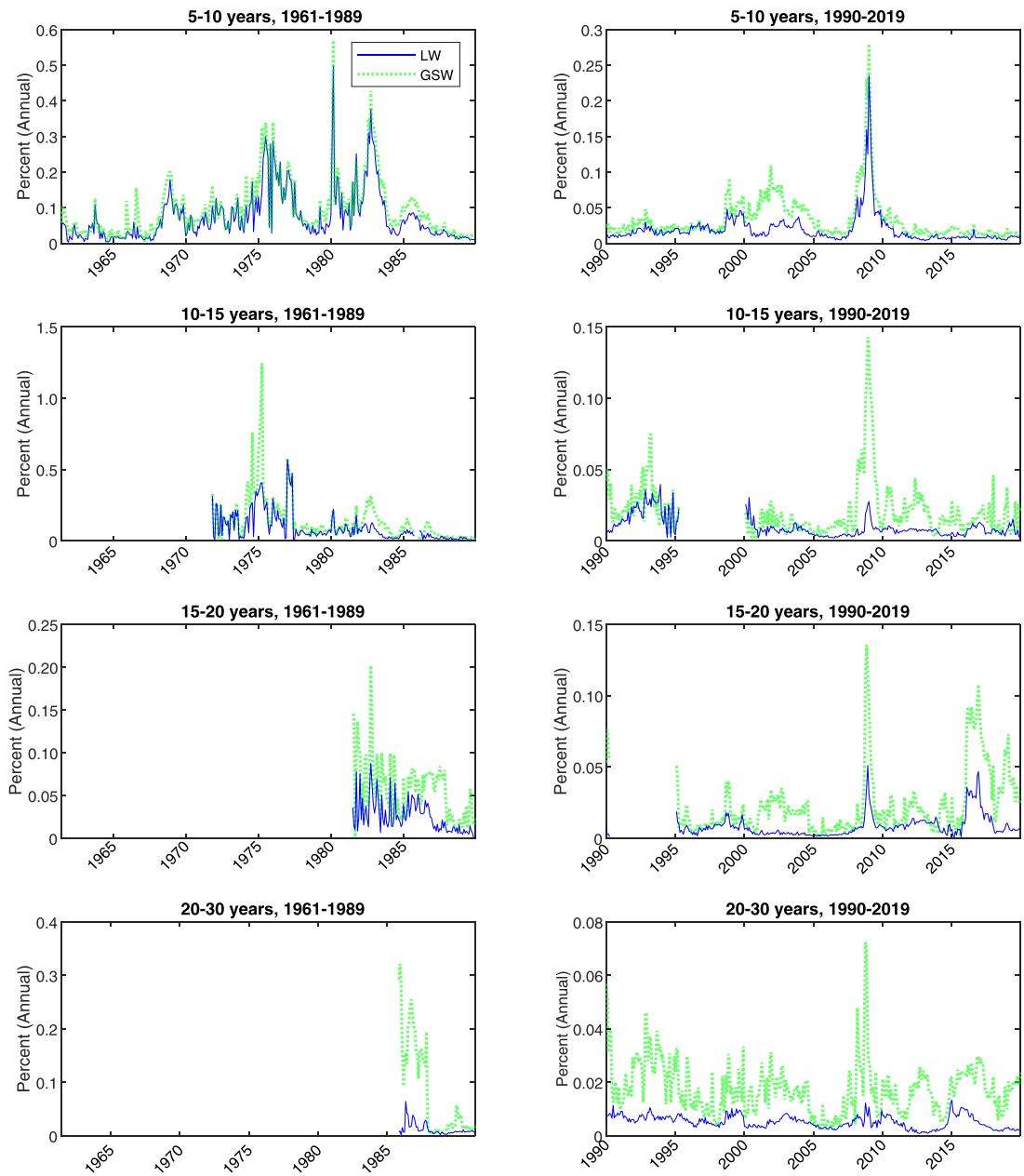


Fig. B.3. Time series of mean absolute error in YTM: The long end. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW, over 1961–1989 (left panels) and 1990–2019 (right panels). We group bonds into four maturity buckets: 5–10 years (top row), 10–15 years (second row), 15–20 years (third row), and 20–30 years (bottom row).

B.2. Out of sample

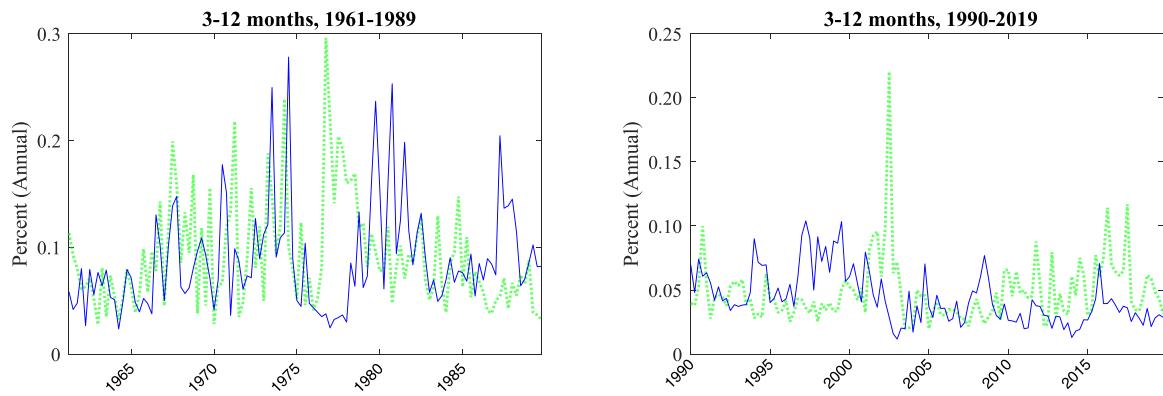


Fig. B.4. Time series of mean absolute error in YTM: Between three months and one year, leave-one-out. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over 1961–1989 (left panels) and 1990–2019 (right panels). We calculate the mean absolute pricing error across maturities between three months and one year. We report out-of-sample pricing errors at the quarterly frequency.

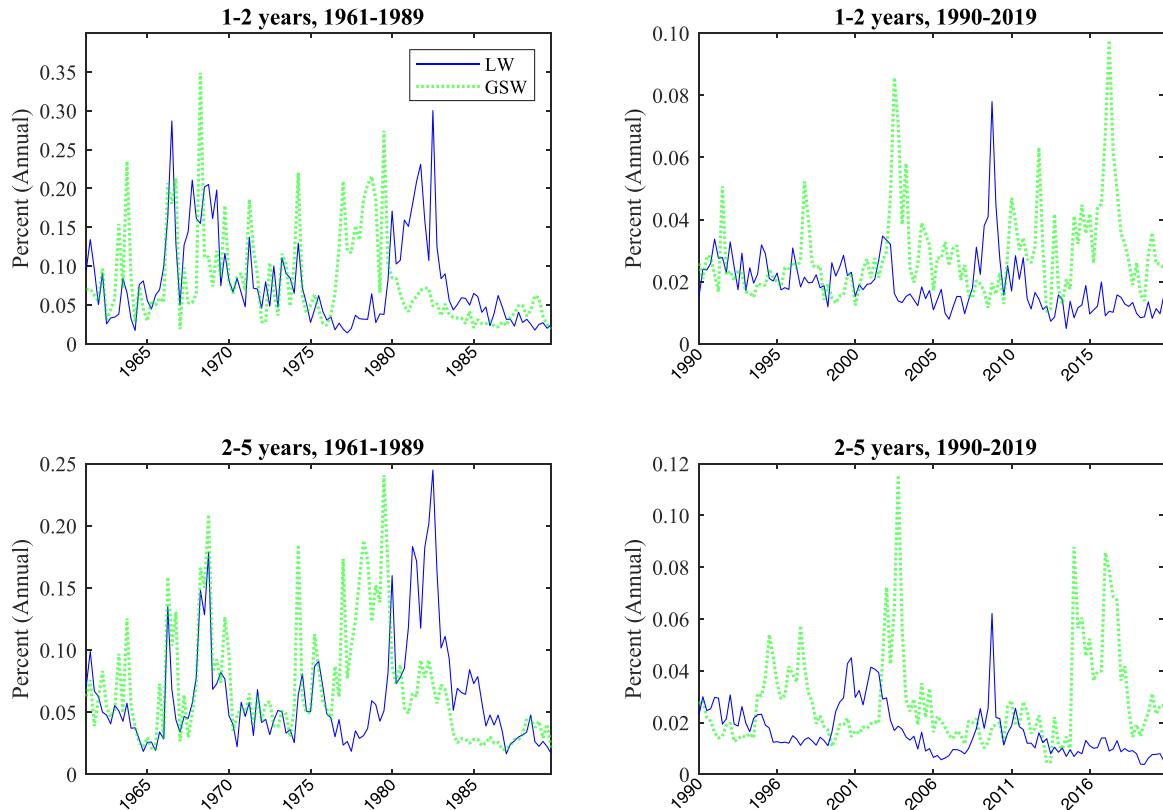


Fig. B.5. Time series of mean absolute error in YTM: The medium end, leave-one-out. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW over 1961–1989 (left panels) and 1990–2019 (right panels). We group bonds into two maturity buckets: 1–2 years (top row) and 2–5 years (bottom row). We report out-of-sample pricing errors at the quarterly frequency.

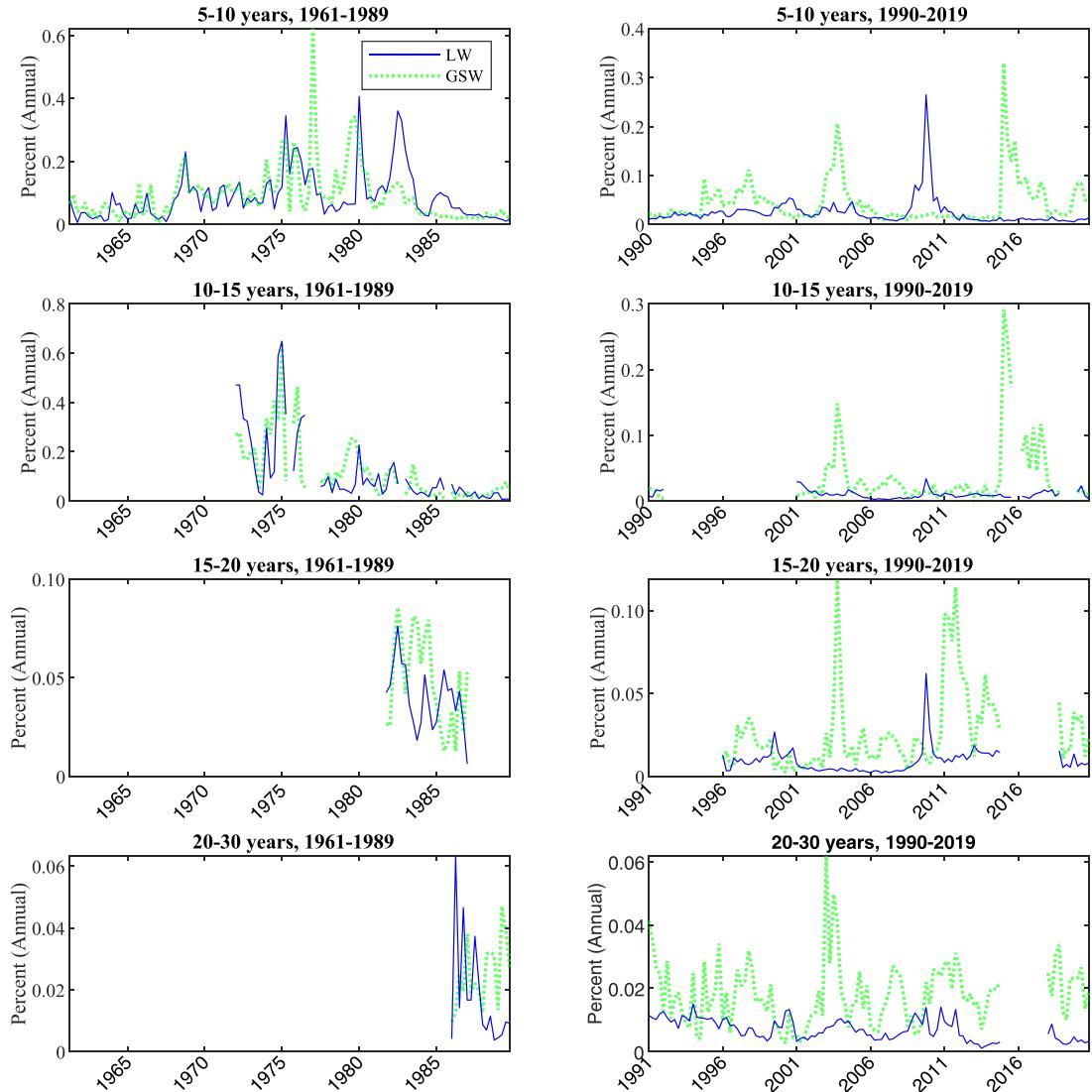


Fig. B.6. Time series of mean absolute error in YTM: The long end, leave-one-out. We plot the mean absolute pricing errors in YTM (i.e., MAYE) for our method and GSW, over 1961–1989 (left panels) and 1990–2019 (right panels). We group bonds into four maturity buckets: 5–10 years (top row), 10–15 years (second row), 15–20 years (third row), and 20–30 years (bottom row). We report out-of-sample pricing errors at the quarterly frequency.

Appendix C. Robustness

C.1. Alternative specifications of our model

Table C.1

Alternative weighting of short end. We present results for one alternative specification of our model. We under-weight pricing errors for maturities below one year by 50% (relative to our baseline specification). For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root mean-squared pricing error (RMSPE), duration-weighted root mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the average pricing errors from June 1961 to December 2019.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW: Alternative Weighting of Short End										
RMSPE	0.016	0.046	0.070	0.133	0.226	0.384	0.380	0.183	0.116	0.157
WRMSPE	0.013	0.040	0.069	0.125	0.224	0.379	0.378	0.184	0.116	0.052
MAPE	0.013	0.035	0.055	0.101	0.189	0.331	0.335	0.154	0.090	0.084
WMAPE	0.010	0.030	0.054	0.094	0.187	0.328	0.334	0.155	0.091	0.022
MAPE (Bliss)	0.006	0.018	0.026	0.056	0.130	0.251	0.240	0.108	0.057	0.052
WMAPE (Bliss)	0.005	0.015	0.026	0.051	0.128	0.248	0.239	0.108	0.058	0.012
MAYE	0.112	0.067	0.040	0.033	0.038	0.052	0.042	0.012	0.006	0.055
HR (Bliss)	0.390	0.383	0.537	0.404	0.314	0.201	0.293	0.316	0.356	0.402
GSW										
RMSPE	0.041	0.058	0.086	0.173	0.307	0.531	0.625	0.440	0.460	0.246
WRMSPE	0.033	0.053	0.085	0.163	0.305	0.524	0.619	0.442	0.457	0.082
MAPE	0.036	0.046	0.069	0.132	0.265	0.465	0.555	0.399	0.402	0.135
WMAPE	0.028	0.043	0.068	0.124	0.262	0.457	0.548	0.402	0.400	0.041
MAPE (Bliss)	0.028	0.030	0.037	0.085	0.203	0.379	0.451	0.346	0.363	0.101
WMAPE (Bliss)	0.022	0.028	0.037	0.078	0.200	0.371	0.445	0.348	0.361	0.030
MAYE	0.336	0.097	0.051	0.044	0.053	0.071	0.068	0.030	0.028	0.115
HR (Bliss)	0.230	0.341	0.412	0.294	0.184	0.118	0.146	0.114	0.124	0.282

Table C.2

Alternative threshold for detecting outliers. We present results for one alternative specification of our model. We use $2 \times IQR$ instead of $3 \times IQR$ for outlier detection, thereby potentially dropping more outliers. For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root mean-squared pricing error (RMSPE), duration-weighted root mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the average pricing errors from June 1961 to December 2019.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW: Alternative Threshold for Detecting Outliers										
RMSPE	0.014	0.041	0.079	0.141	0.246	0.410	0.395	0.192	0.116	0.167
WRMSPE	0.011	0.036	0.077	0.131	0.244	0.405	0.394	0.193	0.116	0.056
MAPE	0.011	0.031	0.061	0.104	0.201	0.349	0.347	0.160	0.090	0.087
WMAPE	0.009	0.027	0.059	0.097	0.200	0.345	0.346	0.161	0.091	0.023
MAPE (Bliss)	0.005	0.015	0.031	0.060	0.141	0.266	0.249	0.113	0.057	0.054
WMAPE (Bliss)	0.004	0.013	0.030	0.055	0.139	0.262	0.248	0.114	0.058	0.012
MAYE	0.084	0.060	0.044	0.035	0.042	0.055	0.043	0.013	0.006	0.049
HR (Bliss)	0.438	0.414	0.518	0.398	0.297	0.191	0.277	0.313	0.357	0.416
GSW										
RMSPE	0.041	0.058	0.086	0.173	0.307	0.531	0.625	0.440	0.460	0.246
WRMSPE	0.033	0.053	0.085	0.163	0.305	0.524	0.619	0.442	0.457	0.082
MAPE	0.036	0.046	0.069	0.132	0.265	0.465	0.555	0.399	0.402	0.135
WMAPE	0.028	0.043	0.068	0.124	0.262	0.457	0.548	0.402	0.400	0.041
MAPE (Bliss)	0.028	0.030	0.037	0.085	0.203	0.379	0.451	0.346	0.363	0.101
WMAPE (Bliss)	0.022	0.028	0.037	0.078	0.200	0.371	0.445	0.348	0.361	0.030
MAYE	0.336	0.097	0.051	0.044	0.053	0.071	0.068	0.030	0.028	0.115
HR (Bliss)	0.230	0.341	0.412	0.294	0.184	0.118	0.146	0.114	0.124	0.282

Table C.3

Alternative values of N_0 : Out-of-sample. We present out-of-sample results for two models: LW at $N_0 = 4$ and LW at $N_0 = 12$. For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root mean-squared pricing error (RMSPE), duration-weighted root mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the averaged pricing errors over the full sample from June 1961 to December 2019 at the quarterly frequency.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW with $N_0 = 4$										
RMSPE	0.015	0.045	0.082	0.157	0.290	0.474	0.550	0.195	0.135	0.197
WRMSPE	0.013	0.039	0.080	0.147	0.287	0.468	0.550	0.196	0.135	0.062
MAPE	0.012	0.033	0.064	0.118	0.242	0.405	0.492	0.160	0.103	0.098
WMAPE	0.009	0.029	0.063	0.110	0.239	0.400	0.491	0.161	0.103	0.024
MAPE (Bliss)	0.005	0.017	0.034	0.073	0.180	0.321	0.391	0.112	0.069	0.065
WMAPE (Bliss)	0.004	0.014	0.033	0.066	0.177	0.316	0.390	0.113	0.069	0.013
MAYE	0.097	0.064	0.047	0.039	0.049	0.063	0.061	0.013	0.007	0.054
HR (Bliss)	0.459	0.401	0.505	0.371	0.265	0.172	0.290	0.296	0.326	0.406
LW with $N_0 = 12$										
RMSPE	0.016	0.047	0.085	0.160	0.288	0.490	0.562	0.277	0.194	0.204
WRMSPE	0.014	0.041	0.083	0.150	0.285	0.484	0.560	0.279	0.194	0.064
MAPE	0.013	0.036	0.068	0.119	0.241	0.419	0.507	0.237	0.152	0.105
WMAPE	0.011	0.031	0.066	0.112	0.238	0.413	0.505	0.238	0.152	0.026
MAPE (Bliss)	0.006	0.019	0.036	0.073	0.178	0.335	0.400	0.188	0.116	0.071
WMAPE (Bliss)	0.005	0.016	0.035	0.067	0.175	0.330	0.398	0.189	0.116	0.014
MAYE	0.113	0.069	0.050	0.040	0.049	0.066	0.063	0.018	0.010	0.059
HR (Bliss)	0.398	0.370	0.478	0.359	0.239	0.173	0.188	0.234	0.229	0.366

C.2. Alternative specifications of GSW

Table C.4

Alternative specifications of GSW. We present results for two models: LW (our model) and GSW. We consider three versions of GSW: "GSW" (we re-estimate GSW using our data but filter out the short end as described in their paper), "GSW, Published Parameters" (we use the parameter values published online: <https://www.federalreserve.gov/data/nominal-yield-curve.htm>), and "GSW with Short-End" (we re-estimate GSW based on our data, including the short end). For each maturity bucket (or across all bonds) and for each date, we calculate eight measures of pricing error: root mean-squared pricing error (RMSPE), duration-weighted root mean-squared pricing error (WRMSPE), mean absolute pricing error (MAPE), duration-weighted absolute pricing error (WMAPE), mean absolute pricing error adjusted for bid-ask spread (MAPE (Bliss)), duration-weighted absolute pricing error adjusted for bid-ask spread (WMAPE (Bliss)), mean absolute yield error (MAYE), and the hit rate (HR (Bliss)). RMSPE, WRMSPE, MAPE, WMAPE, MAPE (Bliss), and WMAPE (Bliss) are based on a face value of \$100. MAYE is based on annualized percentage yield. We report the averaged pricing errors over the full sample from June 1961 to December 2019.

	Maturity Bucket									All
	[0,3mth]	[3mth, 1yr]	[1yr, 2yr]	[2yr,5r]	[5yr, 7yr]	[7yr, 10yr]	[10yr, 15yr]	[15yr, 20yr]	[20yr, 30yr]	
LW										
RMSPE	0.015	0.044	0.073	0.133	0.227	0.391	0.388	0.183	0.116	0.158
WRMSPE	0.012	0.038	0.071	0.125	0.225	0.387	0.387	0.184	0.116	0.051
MAPE	0.012	0.033	0.057	0.101	0.189	0.337	0.343	0.154	0.090	0.084
WMAPE	0.009	0.029	0.056	0.094	0.187	0.333	0.341	0.154	0.091	0.022
MAPE (Bliss)	0.005	0.016	0.028	0.057	0.130	0.257	0.248	0.107	0.057	0.052
WMAPE (Bliss)	0.004	0.014	0.027	0.052	0.128	0.253	0.246	0.108	0.058	0.011
MAYE	0.101	0.063	0.042	0.033	0.038	0.053	0.042	0.012	0.006	0.052
HR (Bliss)	0.424	0.402	0.524	0.403	0.313	0.197	0.290	0.317	0.357	0.412
GSW										
RMSPE	0.041	0.058	0.086	0.173	0.307	0.531	0.625	0.440	0.460	0.246
WRMSPE	0.033	0.053	0.085	0.163	0.305	0.524	0.619	0.442	0.457	0.082
MAPE	0.036	0.046	0.069	0.132	0.265	0.465	0.555	0.399	0.402	0.135
WMAPE	0.028	0.043	0.068	0.124	0.262	0.457	0.548	0.402	0.400	0.041
MAPE (Bliss)	0.028	0.030	0.037	0.085	0.203	0.379	0.451	0.346	0.363	0.101
WMAPE (Bliss)	0.022	0.028	0.037	0.078	0.200	0.371	0.445	0.348	0.361	0.030
MAYE	0.336	0.097	0.051	0.044	0.053	0.071	0.068	0.030	0.028	0.115
HR (Bliss)	0.230	0.341	0.412	0.294	0.184	0.118	0.146	0.114	0.124	0.282
GSW, Published Parameters										
RMSPE	0.054	0.070	0.098	0.189	0.307	0.590	0.685	0.435	0.513	0.271
WRMSPE	0.044	0.067	0.096	0.179	0.305	0.580	0.679	0.437	0.509	0.094
MAPE	0.049	0.057	0.080	0.146	0.259	0.513	0.604	0.393	0.450	0.149
WMAPE	0.037	0.056	0.079	0.138	0.256	0.504	0.598	0.394	0.446	0.051
MAPE (Bliss)	0.040	0.039	0.046	0.096	0.196	0.427	0.499	0.339	0.410	0.114
WMAPE (Bliss)	0.030	0.039	0.045	0.090	0.194	0.418	0.493	0.341	0.407	0.039
MAYE	0.443	0.124	0.059	0.049	0.052	0.079	0.075	0.030	0.031	0.144
HR (Bliss)	0.180	0.264	0.362	0.260	0.194	0.111	0.123	0.114	0.094	0.240

References

- Andreasen, M.M., Christensen, J.H., Rudebusch, G.D., 2019. Term structure analysis with big data: one-step estimation using bond prices. *J. Econ.* 212 (1), 26–46.
- Ang, A., Piazzesi, M., 2003. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *J. Monet. Econ.* 50, 745–787.
- Bauer, M.D., Hamilton, J.D., 2018. Robust bond risk premia. *Rev. Financ. Stud.* 31 (2), 399–448.
- Bauer, M.D., Rudebusch, G.D., Wu, J.C., 2012. Correcting estimation bias in dynamic term structure models. *J. Bus. Econ. Stat.* 30 (3), 454–467.
- Bauer, M.D., Rudebusch, G.D., Wu, J.C., 2014. Term premia and inflation uncertainty: empirical evidence from an international panel dataset: comment. *Am. Econ. Rev.* 104 (1), 323–337.
- Bernanke, B.S., Reinhart, V.R., 2004. Conducting monetary policy at very low short-term interest rates. *Am. Econ. Rev.* 94 (2), 85–90.
- Bliss, R.R., 1996. Testing term structure estimation methods. Technical Report. Working Paper, Federal Reserve Bank of Atlanta.
- Chernov, M., Creal, D., 2020. The ppp view of multi-horizon currency risk premiums. Working paper, UCLA Anderson School of Management.
- Chernov, M., Mueller, P., 2012. The term structure of inflation expectations. *J. Financ. Econ.* 106 (2), 367–394.
- Cieslak, A., 2018. Short-rate expectations and unexpected returns in treasury bonds. *Rev. Financ. Stud.* 31 (9), 3265–3306.
- Cieslak, A., Povala, P., 2015. Expected returns in treasury bonds. *Rev. Financ. Stud.* 28 (10), 2859–2901.
- Cochrane, J. H., 2015. Comments on robust bond risk premia by michael bauer and jim hamilton. Unpublished working paper. University of Chicago.
- Cochrane, J.H., Piazzesi, M., 2005. Bond risk premia. *Am. Econ. Rev.* 95 (1), 138–160.
- Cochrane, J. H., Piazzesi, M., 2009. Decomposing the yield curve. Unpublished working paper. University of Chicago.
- Cooper, I., Priestley, R., 2009. Time-varying risk premiums and the output gap. *Rev. Financ. Stud.* 22 (7), 2801–2833.
- Creal, D.D., Wu, J.C., 2015. Estimation of affine term structure models with spanned or unspanned stochastic volatility. *J. Econ.* 185 (1), 60–81.
- Creal, D.D., Wu, J.C., 2020. Bond risk premia in consumption-based models. *Quant. Econ.* forthcoming.
- Diebold, F.X., Rudebusch, G.D., 2013. Yield Curve Modeling and Forecasting. Princeton University Press, Princeton, NJ.
- Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. *J. Finan.* 57 (1), 405–443.
- Duffee, G.R., 2011. Information in (and not in) the term structure. *Rev. Financ. Stud.* 24 (9), 2895–2934.
- Fama, E., Bliss, R.R., 1987. The information in long-maturity forward rates. *Am. Econ. Rev.* 77 (4), 680–692.
- Fama, E.F., 1984. Term premiums in bond returns. *J. Financ. Econ.* 13 (4), 529–546.
- Fan, J., Gijbels, I., 1995. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. Ser. B (Methodol.)* 371–394.
- Fontaine, J.-S., Garcia, R., 2012. Bond liquidity premia. *Rev. Financ. Stud.* 25 (4), 1207–1254.
- Giglio, S., Kelly, B., 2018. Excess volatility: beyond discount rates. *Q. J. Econ.* 133 (1), 71–127.
- Gilchrist, S., Zakrajšek, E., 2012. Credit spreads and business cycle fluctuations. *Am. Econ. Rev.* 102 (4), 1692–1720.
- Greenwood, R., Vayanos, D., 2014. Bond supply and excess bond returns. *Rev. Financ. Stud.* 27 (3), 663–713.
- Gürkaynak, R.S., Sack, B., Swanson, E., 2005. The sensitivity of long-term interest rates to economic news: evidence and implications for macroeconomic models. *Am. Econ. Rev.* 95 (1), 425–436.
- Gürkaynak, R.S., Sack, B., Wright, J.H., 2007. The u.s. treasury yield curve: 1961 to the present. *J. Monet. Econ.* 54 (8), 2291–2304.
- Gürkaynak, R.S., Sack, B., Wright, J.H., 2010. The tips yield curve and inflation compensation. *Am. Econ. J. Macroecon.* 2 (1), 70–92.
- Hamilton, J.D., Wu, J.C., 2012. Identification and estimation of gaussian affine term structure models. *J. Econ.* 168 (2), 315–331.
- Hamilton, J.D., Wu, J.C., 2014. Testable implications of affine term structure models. *J. Econ.* 178, 231–242.
- Hanson, S.G., Stein, J.C., 2015. Monetary policy and long-term real rates. *J. Financ. Econ.* 115 (3), 429–448.
- Hodrick, R.J., Tomunen, T., 2018. Taking the Cochrane-Piazzesi Term Structure Model Out of Sample: More Data, Additional Currencies, and FX Implications. Working Paper. National Bureau of Economic Research doi:10.3386/w25092.
- Hull, J., White, A., 1990. Pricing interest-rate-derivative securities. *Rev. Financ. Stud.* 3 (4), 573–592.
- Jarrow, R., Yildirim, Y., 2003. Pricing treasury inflation protected securities and related derivatives using an HJM model. *J. Financ. Quant. Anal.* 38 (2), 337–358.
- Jeffrey, A., Linton, O., Nguyen, T., 2006. Flexible term structure estimation: which method is preferred? *Metrica* 63 (1), 99–122. doi:10.1007/s00184-005-0017-8.
- Joslin, S., Priebsch, M., Singleton, K.J., 2014. Risk premiums in dynamic term structure models with unspanned macro risks. *J. Finance* 69 (3), 1197–1233.
- Joslin, S., Singleton, K.J., Zhu, H., 2011. A new perspective on gaussian dynamic term structure models. *Rev. Financ. Stud.* 27, 926–970.
- Linton, O., Mammen, E., Nielsen, J.P., Tanggaard, C., 2001. Yield curve estimation by kernel smoothing methods. *J. Econ.* 105 (1), 185–223.
- Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. *Rev. Financ. Stud.* 22 (12), 5027–5067.
- Lustig, H., Stathopoulos, A., Verdelhan, A., 2019. The term structure of currency carry trade risk premia.. *Am. Econ. Rev.* forthcoming
- Nelson, C.R., Siegel, A., 1987. Parsimonious modelling of yield curves. *J. Bus.* 60 (4), 473–489.
- Pancost, N. A., 2020. Zero-coupon yields and the cross-section of bond prices. Available at SSRN 3157271.
- Park, B.U., Marron, J.S., 1990. Comparison of data-driven bandwidth selectors. *J. Am. Stat. Assoc.* 85 (409), 66–72.
- Rudebusch, G.D., Swanson, E.T., 2012. The bond premium in a DSGE model with long-run real and nominal risks.. *Am. Econ. J. Macroecon.* 4 (1), 105–143.
- Ruppert, D., Sheather, S.J., Wand, M.P., 1995. An effective bandwidth selector for local least squares regression. *J. Am. Stat. Assoc.* 90 (432), 1257–1270.
- Svensson, L. E., 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992 – 1994. Working Paper 4871. National Bureau of Economic Research. 10.3386/w4871
- Swanson, E.T., Williams, J.C., 2014. Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *Am. Econ. Rev.* 104 (10), 3154–3185(32)
- Tukey, J.W., 1977. Exploratory Data Analysis, 2. Reading, Mass..
- Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. CRC Press.
- Wu, J.C., Xia, F.D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. *J. Money Credit Bank* 48 (2–3), 253–291.